# GreatTurbo Cluster Server 10

# 用户手册

# 目 录

前言........1
GreatTurbo Cluster Server 10 的历史........1
本用户手册使用说明........1
第一章　GreatTurbo Cluster Server10 介绍........3
1.1 GreatTurbo Cluster Server 10 简介........3
1.2 应用支持........4
1.3 GreatTurbo Cluster Server 10 技术特点........4
1.3.1 支持磁盘镜像功能........4
1.3.2 提供多种类型的磁盘阵列支持........5
1.3.3 多种硬件心跳保证系统一致性........5
1.3.4 可靠的故障时切换策略........5
1.3.5 支持 STONITH 技术........6
1.3.6 支持 Watchdog Timers........6
1.3.7 智能的服务回迁以及多服务的负载分担........6
1.3.8 可以检测更多的故障........6
1.3.9 应用程序代理检查........7
1.3.10 应用程序代理 API........7
1.3.11 图形管理工具........7
1.3.12 更好的日志文件系统支持........8
1.3.13 更详细的系统故障日志信息........8
1.4 GreatTurbo Cluster Server 10 的使用限制........8
1.5 GreatTurbo Cluster Server 10 的相关术语........9
第二章　硬件安装和操作系统的配置........13
2.1 选择硬件配置........13
2.1.1 集群硬件表........17
2.1.2 最低集群配置举例........23
2.1.3 无单点故障配置举例........25
2.2 集群系统的设置步骤........27
2.2.1 安装基本系统硬件........28
2.2.2 安装控制台交换机........30
2.2.3 安装网络交换机或集线器........31
2.3 安装和配置 Linux 发行版 OS........31
2.3.1 安装 GTES10 操作系统........32
2.3.2 编辑/etc/hosts 文件........33
2.3.3 显示控制台启动信息........34
2.3.4 确定内核中已配置的设备........37
2.4 安装和连接集群硬件的步骤........38
2.4.1 配置心跳通道........39
2.4.2 配置电子开关........40
2.4.3 配置 UPS 系统........42
2.4.4 配置共享磁盘存储器........44

2.4.5 配置磁盘镜像存储器 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
第三章　安装 GreatTurbo Cluster Server 10 ..............................................58
3.1 安装 GreatTurbo Cluster Server 10 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 58
3.1.1 确认您所使用的 GreatTurbo Cluster Server 10 产品的类型 . . 58
3.1.2 安装 GreatTurbo Cluster Server 10 . . . . . . . . . . . . . . . . . . . . . . . 58
3.1.3 安装 guiadmin 客户端 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
3.2 注册 GreatTurbo Cluster Server 10 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
3.3 升级 GreatTurbo Cluster Server 10 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64
3.3.1 GreatTurbo Cluster server 10 的升级 . . . . . . . . . . . . . . . . . . . . 64
第四章　卸载 GreatTurbo Cluster Server 10 ..............................................66
4.1 卸载 GreatTurbo Cluster Server 10 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
4.2 卸载 drbd . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
4.3 卸载 GreatTurbo Cluster Server10 realserver 包 . . . . . . . . . . . . . . . . 68
4.4 卸载 guiadmin 客户端 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69
第五章　配置 GreatTurbo Cluster Server 10 ..............................................70
5.1 member_config 工具说明 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
5.2 配置 GreatTurbo Cluster Server 10 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71
5.2.1 选择其中一节点进行初始化配置 . . . . . . . . . . . . . . . . . . . . . . . . . 71
5.2.2 在对方节点上同步配置 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
5.2.3 利用备份的配置文件配置 GreatTurbo Cluster Server 10 . . . . 81
5.3 运行 GreatTurbo Cluster Server 10 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 81
5.4 停止 GreatTurbo Cluster Server 10 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82
第六章　配置和管理工具说明 ...........................................................................84
6.1 cluadmin 工具 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
6.2 guiadmin 工具 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91
6.2.1 guiadmin 介绍 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91
6.2.2 guiadmin 模块介绍 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 92
6.2.3 登陆密码的设定 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94
第七章　配置和管理 GreatTurbo Cluster Server 10 的服务 ...........................96
7.1 配置服务前的准备工作 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
7.1.1 收集服务信息 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
7.1.2 创建服务脚本 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 98
7.1.3 应用代理 API . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103
7.1.4 配置服务的磁盘存储设备 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 108
7.2 用文本工具 cluadmin 配置和管理服务 . . . . . . . . . . . . . . . . . . . . . . . . 118
7.2.1 添加服务 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 118
7.2.2 显示服务配置 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 125
7.2.3 修改服务 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
7.2.4 删除服务 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128
7.2.5 禁用服务 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128
7.2.6 启用服务 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 129
7.2.7 切换服务 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 130
7.3 用图形工具 guiadmin 配置和管理服务 . . . . . . . . . . . . . . . . . . . . . . . . 130
7.3.1 添加服务 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 132

7.3.2 显示服务配置 ........138
7.3.3 修改服务 ........138
7.3.4 删除服务 ........138
7.3.5 禁用服务 ........138
7.3.6 启用服务 ........139
7.3.7 切换服务 ........139
7.4 配置典型应用的服务........139
7.4.1 配置 Oracle 服务 ........140
7.4.2 配置 MySQL 服务 ........150
7.4.3 配置 DB2 服务 ........158
7.4.4 配置 Apache 服务 ........167
7.4.5 配置 Domino 服务 ........176
7.4.6 配置 Informix 服务 ........182
7.4.7 配置 Sybase 服务 ........188
7.4.8 配置 Websphere 服务........192
7.5 处理错误状态下的服务........197
第八章 集群管理........200
8.1 使用文本界面进行集群管理........200
8.1.1 显示集群和服务的状态 ........200
8.1.2 启动和停止集群软件 ........205
8.1.3 修改集群配置 ........206
8.1.4 备份和恢复集群数据库 ........206
8.1.5 修改集群事件日志 ........207
8.1.6 更新集群软件 ........209
8.1.7 重新加载集群数据库 ........210
8.1.8 修改集群名称 ........211
8.1.9 重新初始化集群 ........211
8.1.10 删除集群中的成员 ........212
8.1.11 修改集群 watchdog 的超时时间 ........213
8.1.12 修改集群的心跳属性........213
8.1.13 修改集群告警邮件属性 ........215
8.2 使用图形界面进行集群管理........216
8.2.1 显示集群和服务的状态 ........216
8.2.2 如何在集群系统上运行 guiadmin ........217
8.2.3 如何从远程系统上运行 guiadmin ........217
8.2.4 配置、修改心跳（heartbeat）参数........218
8.2.5 配置、修改集群进程日志级别 ........219
8.2.6 配置、修改邮件提示参数 ........219
8.2.7 配置、修改 watchdog 信息 ........220
8.2.8 显示节点的配置信息 ........221
第九章 集群的维护........222
9.1 GreatTurbo Cluster Server 10 的日志信息 ........222
9.2 Log 收集工具的使用方法 ........223
9.3 FAQ ........224

9.4 诊断和纠正集群中的故障 ........................................228
9.5 联系拓林思软件有限公司 ........................................231
附录 A 补充硬件信息 ........................................233
A.1 设置 Cyclades 终端服务器 ........................................233
A.1.1 设置路由器的 IP 地址 ........................................234
A.1.2 设置网络和终端端口参数 ........................................236
A.1.3 配置 Turbolinux，以向控制台端口发送控制台信息 ........................................241
A.1.4 连接到控制台端口 ........................................243
A.2 设置 RPS-10 电子开关 ........................................244
A.3 SCSI 总线配置要求 ........................................245
A.3.1 SCSI 总线终端 ........................................247
A.3.2 SCSI 总线长度 ........................................249
A.3.3 SCSI 标识号 ........................................249
附录 B 补充软件信息 ........................................251
B.1 集群通信机制 ........................................251
B.2 集群守护进程（Daemon） ........................................252
B.3 故障切换和恢复情形 ........................................253
B.3.1 系统挂起（Hang） ........................................253
B.3.2 系统紧急（Panic） ........................................254
B.3.4 网络连接完全故障 ........................................254
B.3.5 远程电子开关连接故障 ........................................255
B.3.6 集群 Daemon 故障 ........................................255
B.4 集群数据库选项 ........................................255
B.5 磁盘镜像配置 ........................................257
B.5.1 介绍 ........................................258
B.5.2 兼容性 ........................................259
B.5.3 Linux kernel2.4 版本 drbd 的配置文件 ........................................259
B.5.4 Linux kernel2.6 版本 drbd 的配置文件 ........................................261
B.5.5 drbd 的启动和停止 ........................................264
B.5.6 Linux kernel2.4 版本/proc/drbd ........................................264
B.5.7 Linux kernel2.6 版本/proc/drbd ........................................265
B.5.8 有关 drbd 的 Q&A ........................................267
B.6 调整 Oracle 服务 ........................................268

# 前言

## GreatTurbo Cluster Server 10 的历史

GreatTurbo Cluster Server 10 是北京拓林思软件有限公司推出的、为满足 Linux 平台电信级和企业级应用的高可用产品。GreatTurbo Cluster Server 10 提供的双机高可用、负载均衡方案能够更好的满足用户业务的连续性，并且能够满足不同应用对高可用的要求。

自 2001 年以来，北京拓林思软件有限公司先后推出了高可用和负载均衡产品。GreatTurbo Cluster Server 10 是完整的高可用、负载均衡解决方案。 GreatTurbo Cluster Server 10 提供更好的可靠性、更高的性价比、更好的易用性和可管理性，完全满足企业级应用所要求的 RASM(Reliability、Availability、Scalability、Manageability)特性。同时 GreatTurbo Cluster Server 10 支持磁盘镜像功能，使得用户应用和 GreatTurbo Cluster Server 10 软件可以不需要磁盘阵列来提供高可用服务。

GreatTurbo Cluster Server 10 可支持 Turbolinux 发布的 GTES10 操作系统平台。由于 GreatTurbo Cluster Server 10 可以和 Turbolinux 操作系统更好的配合，使得从操作系统到 GreatTurbo Cluster Server 10 软件构建的高可用解决方案更加的可靠，并广泛服务于电信、银行、政府等行业客户。

## 本用户手册使用说明

《GreatTurbo Cluster Server 10 用户手册》是帮助用户了解、使用和维护 GreatTurbo Cluster Server 10 的文档，文档中详细阐述了 GreatTurbo Cluster Server 10 的技术特点、系统软硬件环境、安装、配置、管理、维护、FAQ 等各个方面。

如果用户第一次使用 GreatTurbo Cluster Server 10，建议仔细阅读本手册，以便于对 GreatTurbo Cluster Server 10 进行详细的了解。

如果用户已经对 GreatTurbo Cluster Server 10 有所了解，可以直接阅读

《GreatTurbo Cluster Server 10 快速安装文档》进行安装配置。

# 第一章 GreatTurbo Cluster Server10 介绍

## 1.1 GreatTurbo Cluster Server 10 简介

GreatTurbo Cluster Server 10 是北京拓林思软件有限公司推出的为满足 Linux 平台企业级应用的高可用和负载均衡产品。它同时具备了高可用和负载均衡产品所有的功能，而且在原有功能的基础上增加了许多新的功能。它提供的双机高可用方案不但能够保证负载均衡调度器自身的高可用，而且更好的满足用户业务的连续性、更加可靠，并且能够满足不同应用对高可用的要求；而其提供的负载均衡功能使得您的 Web、IPTV、Mail、Game 等各种解决方案可以昼夜不停地提供 24×7 的服务，同时提供了您业务强大的并发处理能力。GreatTurbo Cluster Server 10 是基于最新的 linux 2.6 内核，其领先的集群技术为您的业务注入更高的可靠性、稳地性、和强大的扩展能力。

自 2000 年以来，北京拓林思软件有限公司先后推出了高可用产品系列 TurboHA 和负载均衡产品 TurboCluster 。GreatTurbo Cluster Server 10 是在这两个系列产品的基础上，根据市场的实际需求和企业级用户多年实践经验的总结，依据已有成熟架构的基础上而开发的，它能够为 LAMP（Linux、Apache、Mysql、Perl／PHP／Python）架构的应用和企业级用户提供更加可靠和可扩展的基础平台。GreatTurbo Cluster Server 10 提供更好的可靠性和可扩展性，更高的性价比，更好的易用性和可管理性，完全满足企业级应用所要求的 RASM(Reliability，Availability，Scalability，Manageability)特性。

GreatTurbo Cluster Server 10 可支持 Turbolinux 发布的 GTES10/TDS10 操作系统平台，适用于 i386、x86_64、IA64、openpower 等主流的硬件平台。由于 GreatTurbo Cluster Server 10 可以和 Turbolinux 操作系统更好的配合，使得从操作系统到 GreatTurbo Cluster Server 10 软件构建的高可用和负载均衡解决方案更加的可靠，并广泛服务于电信、银行、政府等行业客户。

## 1.2 应用支持

当我们通过硬件（服务器、交换机、电子开关等）和软件（操作系统平台、HA 系统软件、应用软件等）搭建一个高可用群集环境的时候，首先我们需要明确的是，高可用系统软件能否支持和管理我们的应用程序。GreatTurbo Cluster Server 10 高可用功能能够支持绝大多数的 Linux 环境下的应用程序，支持的典型应用程序类型如下：

- 通用的，无需修改的应用程序：GreatTurbo Cluster Server 10支持大多数Linux平台的应用程序，这些应用大多数是能够接受几秒种的停机时间的业务。
- 数据库应用：GreatTurbo Cluster Server 10能够很好的支持各种数据库产品，包括Oracle 8i/9i/10g、MySQL、Sybase和IBM DB2数据库。
- 各种文件服务：GreatTurbo Cluster Server 10能够为各种类型的文件服务提供高可用集群功能，如NFS和SMB/CIFS (使用Samba)。
- 主流的商业应用软件：GreatTurbo Cluster Server 10能够很好的支持主流的商业应用软件，如SAP、Oracle Application Server和Tuxedo。
- 互联网和开放源代码的应用：GreatTurbo Cluster Server 10可以很好的支持各种流行的互联网应用软件和各种开放源代码产品，如Apache、Wu-ftp、VSFTP等。
- 邮件服务软件：如Sendmail和Domino。

## 1.3 GreatTurbo Cluster Server 10 技术特点

### 1.3.1 支持磁盘镜像功能

磁盘镜像功能，是一种不需要磁盘阵列的双机数据共享方案。它的基本原理是通过对两个节点各自的本地磁盘分区进行实时镜像操作，使得这两个本地磁盘对双方节点而言，可以当作一个虚拟的共享磁盘设备来使用，这个虚拟的 RAID-1 级别的共享磁盘设备，能够作为应用的共享设备，既可以当作共享的裸设备来使用，也可以在其上创建各种 Linux 文件系统。GreatTurbo Cluster Server 10 本身提供磁盘镜像功能，使得共享数据的应用不需要磁盘阵列也能够搭建双机高

可用方案。

### 1.3.2 提供多种类型的磁盘阵列支持

对于需要磁盘阵列的一些应用，如数据库应用等，需要专业的硬件磁盘阵列来保证性能。而 GreatTurbo Cluster Server 10 能够支持绝大多数的磁盘阵列设备。

目前业界使用的磁盘阵列时，一般分两种情况：第一种是带独立 RAID 处理器的磁盘阵列，主流厂商的 SCSI 或光纤磁盘阵列都可以适用于 GreatTurbo Cluster Server 10 的要求。第二种是使用主机 RAID 卡和磁盘柜（磁盘柜是指不具备硬件 RAID 处理器的磁盘盒）的方式，这种磁盘设备通过双机的硬件 RAID 卡 clustering 技术和 RAID 的用户接口工具等，也可以满足 GreatTurbo Cluster Server 10 的共享数据存取的需求，常见的这种类型的“磁盘阵列”的典型产品有 IBM EXP400，DELL 220S，HP MSA500-G2 等。

### 1.3.3 多种硬件心跳保证系统一致性

GreatTurbo Cluster Server 10 同时支持直连网线、串口和 raw 磁盘分区的方式来同步 HA 两个节点之间的心跳信息。可同时支持多条直连网线和串口线以及磁盘的 raw 分区作为通道，提供更高可靠性的硬件冗余方式，以保证两个节点之间不会发生 Split-brain 现象。其中 raw 磁盘分区的心跳通道，保证了只要主备节点能够访问共享数据，就不会发生裂脑，从而有效的确保了共享数据的一致性。即使两节点之间的心跳通道都发生故障，GreatTurbo Cluster Server 10 还可以通过配置第三方参考 IP 的方式,保证两个节点系统的一致性。GreatTurbo Cluster Server 10 支持配置多个第三方参考 IP，避免了第三方参考 IP 成为的单一故障点。

### 1.3.4 可靠的故障时切换策略

无论是否配置第三方 IP，主节点所有的网络都发生故障时，仍能够保证服务切换到正常的备节点上，不影响对外正常提供服务。

### 1.3.5 支持 STONITH 技术

Stonith(shut the other node in the head)，就是把故障节点重启，以保证资源被完全释放。Stonith 的方式有两种,一种是通过电子开关(power switch)来重启对方；(另外一种是通过网络发送命令来重启对方。但是后者通常不起作用)。有一些厂家的服务器有类似的管理界面（一种硬件设备），也可以用来作为 stonith 的工具，比如 IBM xServer 的 RSA(Remote Supervisor Adapter)和 Intel 的 IPMI，GreatTurbo Cluster Server 10 也可以支持。

### 1.3.6 支持 Watchdog Timers

GreatTurbo Cluster Server 10 的高可用功能支持三种类型的看门狗定时器为系统提供了一个稳健的 I/O barrier。最简单的就是 Linux 内核自带的通过中断处理来实现的软件 softdog 定时器，它被 GreatTurbo Cluster Server 10 用来控制后台程序的执行。

Linux 内核也支持一种硬件 NMI (non-maskable interrupt) 看门狗，这种硬件看门狗通常需要专门服务器硬件支持 (常用的是主板上的 Intel 810 TCO 芯片组)。NMI 看门狗在没有检测到一个稳定正常的中断发生时就会确发节点服务器的重启动。

最后一种，就是传统的硬件看门狗定时器，它是一种 PCI 设备，在市面上很常见。当 PCI 设备的驱动没有正常的复位时，它就会强迫系统关闭或重启。

### 1.3.7 智能的服务回迁以及多服务的负载分担

GreatTurbo Cluster Server 10 支持优先节点的设置，可以把一些服务设定到指定的优先节点，当优先节点故障时，服务切换到另一个节点，而当优先节点又恢复时，服务会自动迁移到优先节点。这样可以让多个服务分别运行在两个节点上，使得服务的负载可以分担到两个节点上。

### 1.3.8 可以检测更多的故障

GreatTurbo Cluster Server 10 能够检测更多的系统故障，从而增强了高可用性集群所提供的可靠性。

- 系统故障：　　　硬件错误
- 系统紊乱：　　　统软件错误
- 存储不可访问：　存贮错误
- 网络断开：　　　网络错误
- 集群进程故障：　集群软件错误
- 服务故障：　　　服务应用程序错误。

## 1.3.9 应用程序代理检查

GreatTurbo Cluster Server 10 通过使用应用程序代理检查某一服务是否运行。应用程序代理用于定期检查某一服务是正常工作。如果服务没有正常运行，则相应地触发一次切换，使服务在另一节点被恢复。GreatTurbo Cluster Server 10 提供用于常用服务的应用程序代理，对于自身没有应用程序代理程序的服务则可以使用 GreatTurbo Cluster Server 10 提供的接口进行灵活的按需定制。请同时参见本文中的“ 应用程序代理 API ”一节。

## 1.3.10 应用程序代理 API

应用程序代理 API 是一种在应用程序代理或服务检查程序和 GreatTurbo Cluster Server 10 服务进程之间的接口，在 GreatTurbo Cluster Server 10 用户手册中有详细介绍。按照此接口的规范，您可以为您的特定服务编写定制的应用程序代理。编写定制的应用程序代理的好处在于它可以根据您现场实际的负载情况为应用程序提供更精确的服务检查以及更快的切换。

## 1.3.11 图形管理工具

GreatTurbo Cluster Server 10 高可用功能通过提供基于 Java 技术的图形管理工具而改善了集群的可管理性。支持本地和远程的监控和管理，支持 Linux/Windows 客户端的管理。

利用所提供的图形管理工具，可以方便地进行配置更改和状态监测。除了提供图形管理工具外，GreatTurbo Cluster Server 10 还提供有功能同样强大的命令行配置和监控管理工具。

### 1.3.12 更好的日志文件系统支持

GreatTurbo Cluster Server 10 支持与日志文件系统诸如 Reiser 和 Ext3 等的协同工作。这些日志文件系统特别适用于 GreatTurbo Cluster Server 10，因为它们消除了诸如 Ext2 等文件系统中所化费的耗时的文件系统检查从而减少了切换时间。当系统装载时，日志文件系统仅仅要求恢复其日志。当一种日志文件系统被用于共享存贮时，GreatTurbo Cluster Server 10 能够自动地进行确认，跳过不需要的 FSCK 文件系统检查，并立即装载文件系统用于文件系统日志的恢复。

### 1.3.13 更详细的系统故障日志信息

GreatTurbo Cluster Server 10 采用的日志函数和 Linux 的 syslogd 是一样的方式，在两个节点均有记录，每个守护进程都有自己的日志级别，可以在配置文件中指定。每一条记录的信息，包括有时间、日志级别、进程名称、进程 id、消息等内容，这样可以方便用户进行应用故障现场的保护以及故障后的分析定位。同时日志的级别可以动态进行设置调整，以根据实际需要调整输出日志的信息内容。默认情况下，系统已经将日志级别设置成较为详细的信息输出，包括 HA 启动、停止过程，HA 事件触发原因，服务故障原因，服务切换过程，服务手动操作记录等。同时 GreatTurbo Cluster Server 10 还提供日志收集工具，自动收集系统以及 HA 相关信息，以便于更方便的进行故障定位。

## 1.4 GreatTurbo Cluster Server 10 的使用限制

- ✧ GreatTurbo Cluster Server 10 目前只支持 GTES10/TDS10 操作系统平台。
- ✧ GreatTurbo Cluster Server 10 暂不支持并行处理的应用。也就是说不支持同一个应用在两个节点同时并发运行的应用。例如：Oracle 9i RAC。
- ✧ GreatTurbo Cluster Server 10 配置的所有负载均衡(LB)服务都必须采用同一种 ip 负载均衡技术。
- ✧ GreatTurbo Cluster Server 10 配置的所有服务都必须运行在同一调度节点，并且只能对全部服务进行统一操作，如 Enable、Disable、Relocate 等。

✧ GreatTurbo Cluster Server 10 的稳定性需要 OS 提供支撑。如果当操作系统宕机时，可能会出现因 OS 没有彻底释放资源而导致 HA 系统丧失高可用功能。在这种情况下，除非有额外的硬件设备，否则 GreatTurbo Cluster Server 10 并不能够完全保证能够自动恢复操作系统。此时需要用户手工干预操作系统，对崩溃的操作系统进行复位操作。也就是说，当 OS 宕机时（尽管这种可能性很小），如果用户没有以下硬件作为保障，仍然有可能会出现用户业务中断的可能：

1）采用电子开关。

2）服务器节点采用支持硬件 watchdog 功能的主板。

✧ 配置 GreatTurbo Cluster Server 10 的两台计算机节点的心跳方式时，必须保证至少一条心跳通道正常工作。如果两台节点之间的所有心跳通道都发生故障而不能正常连通，有可能会导致 GreatTurbo Cluster Server 10 发生裂脑（split-brain），发生裂脑后，GreatTurbo Cluster Server 10 有可能会导致用户的资源不一致。为了完全杜绝裂脑现象的发生，可以采取如下方法：

1）采用电子开关。

2）使用第三方参考 IP，有关第三方参考 IP 的介绍将在第三章详述。

其中第一种办法是使用硬件的办法，由于电子开关是额外的电子硬件设备，需要用户自行购买，所以实际中采用这种方式并不多；而第二种方式是软件的方法，可以保证 GreatTurbo Cluster Server 10 发生裂脑时，用户的资源不受损失，但是需要用户提供另一个永久性正常工作的参考性 IP 地址。

# 1.5 GreatTurbo Cluster Server 10 的相关术语

<u>节点</u>： 指运行 GreatTurbo Cluster Server 10 程序的计算机。

<u>服务</u>： 也叫做<u>资源组</u>，指用户应用相关的一组资源的集合，包括用户应用的进程资源，磁盘资源，网卡资源，浮动 IP 资源，drbd 镜像资源等。服务可以是其中几种资源或者全部资源的组合。服务也可以为空，即不包括任何资源。通常用户的一个应用与 GreatTurbo Cluster Server 10 的一个服务对应。GreatTurbo Cluster Server 10 的服务分为两种：高可用服务（HA 服务）和负载均衡服务（LB 服务）。

HA 服务针对高可用集群确保高可用的对象。LB 服务是负载均衡系统进行调度的具体应用，例如 Oracle 服务。HA 服务也可以是具体商业应用。在 GreatTurbo Cluster Server 10 的典型应用中，LB 作为 HA 服务出现。也就是说，LB 的高可用性是由 HA 提供、保证的。

负载调度器（load balancer），也成为调度节点，它是整个集群对外面的前端机，负责将客户的请求发送到一组真实服务器上执行，而客户认为服务是来自一个 IP 地址（我们可称之为虚拟 IP 地址）上的。

真实服务器 （real server）：也叫服务器池（server pool），是一组真正执行客户请求的服务器，执行的服务有 WEB、MAIL、FTP 和 DNS 等。用户请求由调度节点分发给真实服务器，由真实服务器来处理用户请求。

主节点 ： 指服务运行所在的节点。

备节点 ： 指完全没有服务运行的节点。如果主节点发生任何故障，服务就会从主节点迁移到备节点。此时的备节点也就转变成主节点，此前的主节点也就转变成备节点。

主备方式 ：常见的主备方式有 Active-Standby、Active-Active。Active-Standby 是指服务仅在一个节点上运行。Active-Active 是指在两个节点都有服务运行。也就是说，Active-Active 方式时都是主节点，也都是备节点，互为主备关系。

优先节点（preferred node）： 指服务将优先运行的节点。当配置一个服务时，可以给这个服务设定优先节点。一般有两种情况建议配置优先节点，一种情况是当两个服务器节点的硬件配置不一样时，应当将服务的优先节点设定为硬件配置较好的节点；另一种情况是当需要配置多个服务且应用的负载较重，而且希望两个服务分别在两个节点上运行以分担负载时，可以给这两个服务分别各设定一个节点作为优先节点。

例如：假设服务 s 配置的优先节点是节点 A，并且将服务 s 在 A 节点启动，当 A 节点发生故障时，服务 s 会迁移到 B 节点，后来当 A 节点恢复之后，服务 s 会因为 A 节点设置成优先节点而自动回迁移到 A 节点。

当然，配置优先节点带来的弊病是，会导致服务多迁移一次。

服务的迁移（relocate）： 是指服务在一个节点发生故障之后，服务先在故障节点进行服务的停止以释放服务的所有资源，然后在另一节点启动服务，使服务继续可用的过程。

服务迁移的时间 ：是指服务不可用的时间，也就是服务正在恢复之中但尚不可用的时间段。

例如：服务在 A 节点发生故障的时刻为 T(a)，服务自动迁移到 B 节点并在 B 节点成功运行的时刻为 T(b),那么服务的迁移时间 T(relocate)=T(b)-T(a)。另一方面，从整个迁移的子过程来看，服务迁移的时间（近似等于）≈服务在 A 节点检测到错误的时间＋服务在 A 节点停止的时间＋服务在 B 节点启动的时间。其中，服务的启动/停止时间是由用户的应用来决定的，一般不能够进行调整，而服务检测到错误的时间可以通过 GreatTurbo Cluster Server 10 的配置参数来进行调整。

服务检测到错误的时间 ＝ 服务检查的间隔时间（check interval） * 服务连续检查到错误的次数（check count）。Check interval 和 check count 参数可以在配置服务的检测脚本时进行指定。这样，只有当 GreatTurbo HA 检测到连续出错次数达到指定的次数后，GreatTurbo Cluster Server 10 才认为服务确实出错。服务确实出错后，会自动触发 GreatTurbo Cluster Server 10 进行服务的迁移。

裂脑（split-brain）： 所谓裂脑，是指 HA 的节点之间彼此失去了联系，但是单个节点的 HA 仍然运行正常。发生裂脑的充要条件是：1. HA 的两个节点之间的所有心跳通道都发生了故障，导致 HA 的两个节点失去了任何联系。2. 两个节点的 HA 软件正在正常运行。

裂脑带来的直接后果是导致两个节点都会各自启动并运行同一个服务，竞争同一个服务的资源，这样有可能会导致资源（尤其是共享数据）被损坏。

**drbd**（磁盘镜像设备） ： 是指对两个节点各自大小相同的本地磁盘分区进行实时镜像操作，使得这两个本地磁盘对双方节点而言，可以当作一个虚拟的共享磁盘阵列来使用，这个虚拟的共享磁盘阵列就叫做 drbd 设备，可以把 drbd 当作普通的磁盘设备来使用。

**Watchdog**（看门狗） ：Watchdog 分为硬件级和软件级两种。硬件级 watchdog 是用来保障操作系统自动恢复的一种手段，如果服务器节点的主板支持 watchdog 功能，那么在 GreatTurbo Cluster Server 10 中可以进行相应的配置，当操作系统

发生死机等不响应情况，主板就会将操作系统自动重启恢复，而无需人工干预，这一点对于“ 24 * 7 ”方式运行的服务器而言非常有用。而软件级的 watchdog 可以用来保证 GreatTurbo Cluster Server 10 程序的健壮性但并不能保障操作系统的自动重启恢复。所以硬件级的 watchdog 更加实用。

# 第二章 硬件安装和操作系统的配置

设置硬件配置和安装 Linux 发行版时需按以下步骤操作：

1. 选择满足您的应用要求和用户要求的集群硬件配置。
2. 设置并连接集群系统和可选的控制台交换机、网络交换机或网络集线器。
3. 在集群系统节点上安装和配置 Turbolinux 操作系统。
4. 设置其余的集群硬件组件，并将其连接到集群系统。

在设置硬件配置和安装 Linux 发行版之后，您就可以安装集群软件了。

## 2.1 选择硬件配置

GreatTurbo HA 10 允许您使用商用软件来设置集群配置，以满足您的应用和用户对性能、可用性和数据完整性的要求。可供选择的硬件范围很广，从低成本的最低配置（只包括运行集群所需的组件）到高端配置（包括冗余心跳通道、硬件 RAID 和电子开关）一应俱全。

无论您使用那种配置，都需要在集群中使用高质量的硬件，因为硬件故障往往是造成系统停机的主要原因。

尽管所有的集群配置都提供一定的可用性，但有些配置能够避免所有的单点故障。此外，所有的集群配置都提供数据完整性，但有些配置可以提供任何故障情形下的数据保护。因此，您必须完全了解自己计算环境的需求和不同硬件配置的数据完整性特性，以选择符合要求的硬件。

在选择集群硬件配置时，需要考虑以下几个方面：

- 您的应用和用户的性能要求

选择一种可提供足够内存、CPU 和 I/O 资源的硬件配置，还应确保该配置可以满足未来工作负载增长的需要。

- 成本限制

您所选择的硬件配置还必须能够满足预算要求。例如，具有多个 I/O 端口的系统通常比扩展能力较低的低端系统更为昂贵。TurbHA 6 支持一整套存储设备，从单磁盘到多端口、独立式 RAID 控制器无所不包。

- 可用性要求

如果您的计算环境需要最高的可用性（如生产环境），那么您可以设置能够防范所有单点故障（包括磁盘、存储互连、心跳通道和电源故障）集群硬件配置。对于那些可容忍可用性出现中断的环境而言（如开发环境），可能不需要这样的保护。请参见配置心跳通道、配置 UPS 系统和配置共享磁盘存储器，了解关于使用冗余硬件实现高可用性的更多信息。

- 在所有故障条件下保持数据完整性。

在集群配置中使用电子开关可确保服务数据能够在各种故障条件下得到保护。电子开关可在故障切换过程中，在重启另一个集群系统的服务之前，先对该集群系统进行加电重启。电子开关可以防止未响应系统（“挂起”）在服务故障切换后恢复响应（“解除挂起”）时可能引起的数据损坏，并对已从另一个集群系统（服务节点）中接收 I/O 的磁盘做 I/O 操作。

如果存储设备支持 SCSI 保留命令，那么集群还可使用 SCSI 保留来提供故障条件下的数据保护。通过使用 SCSI 保留，系统可以阻止另一个系统访问存储器，直至另一个系统重启并进入已知状态为止。

如果未使用电子开关，也不支持 SCSI 保留，那么集群将通过一种“软件重启动”机制来提供数据完整性。“软件重启动”依靠故障系统

对网络信息的响应功能。如果未通过网络收到通知，则不发生故障切换。通过支持无电子开关和无 SCSI 保留型配置，GreatTurbo Cluster Server 10 可以支持所有不同类型的共享存储器。您可以灵活地选择最佳的解决方案，来满足您对价格、数据完整性和可用性方面的要求。

最低硬件配置仅包括集群操作所需的硬件组件，具体如下：

- 两台运行集群服务的服务器
- 提供心跳通道和客户机网络访问的以太网连接

请参见最低集群配置举例查看这种类型的硬件配置。

最低硬件配置是最为经济高效的一种集群配置；但是它会出现多种单点故障。比如，如果网络连接中断，两个集群节点不能互相通信，那么就不能通过网络实现服务。此外，最低配置中也不包括电子开关，后者可以防范任何故障条件下的数据损坏。因此，只有开发环境适合使用最低集群配置。

为提高可用性和防范组件故障，以及确保各种故障条件下的数据完整性，您可以适当扩展最低配置。下表描述了您可以如何提高可用性和确保数据完整性：

| 防范目标 | 采取措施 |
| --- | --- |
| 磁盘故障 | 使用硬件 RAID 在多个磁盘间复制数据 |
| 存储互连故障 | 使用带有多条 SCSI 总线或光纤通道互连的 RAID 阵列配置 |
| RAID 控制器故障 | 使用双 RAID 控制器提供到磁盘数据的冗余访问能力 |
| 心跳通道故障 | 在集群系统间建立点对点以太网或串行连接 |
| 电源故障 | 使用冗余不间断电源（UPS）系统 |
| 所有故障情况下的数据损坏 | 使用电子开关或 SCSI 保留命令 |

**表 1.对策**

无单点故障硬件可在任何故障情况下确保数据的完整性，其配置包括以下组件：

- 两台服务器：运行集群服务
- 系统间的以太网连接：提供心跳通道和客户机网络接入
- 双控制器 RAID 阵列，用以复制服务数据；应支持 SCSI 保留命令，这样即无需使用电子开关。
- 两个电子开关，可使两个集群系统在故障切换过程中互相重启电源。
- 集群系统间的点对点以太网连接，负责提供冗余的以太网心跳通道。
- 集群系统间的点对点以太网连接，负责提供串行心跳通道。
- 两套 UPS 系统，负责提供高可用性电源。

关于此类硬件配置的举例，请参见无单点故障配置举例。

集群硬件配置还包括其它计算环境中常见的可选硬件组件。例如，您还可以添加网络交换机或网络集线器和控制台交换机，前者可以将集群系统连接到网络，后者可简化多系统的管理，使您无需为每个集群系统单独配备监视器、鼠标和键盘。

终端服务器是一类控制台交换机，通过它您可以连接到串行控制台并远程管理多个系统。或者为了经济考虑，您可以使用 KVM（键盘、显示器和鼠标）交换机，在多个系统之间共享一套键盘、监视器和鼠标。KVM 交换机适合于通过图形化用户界面（GUI）执行管理任务的配置情形。

选择集群系统时，要确保它配有所需的 PCI 插槽、网络插槽、和硬件配置所需要的串行口插槽。比如：无单点故障配置需要多个串行口和以太网端口。所选的集群系统最好有两个以上的串行口。更多详情，请参见安装基本系统硬件。

### 2.1.1 集群硬件表

使用以下列表，确认集群配置所需的硬件组件。虽然在某些情况下集群可以和其它产品一起使用，但我们仍要在列表中列出经过集群测试的专用产品。

| 硬件 | 数量 | 说明 | 要求 |
|---|---|---|---|
| 集群系统 | 2 | GreatTurbo Cluster Server 10 支持 IA-32 硬件平台集群系统必须提供足够的集群硬件配置所需的 PCI 插槽、网络插槽和串行口。在每个集群系统中，磁盘设备的名称必须一致，所以推荐使用具有相同 I/O 子系统的系统。另外，推荐使用 CPU 速度为 450 Mhz ，内存为 256 MB 的系统。详见安装基本系统硬件。 | 是 |

**表 2. 集群系统硬件**

| 硬件 | 数量 | 说明 | 要求 |
|---|---|---|---|
| 电子开关 | 2 | 电子开关能让每个集群系统对其它集群系统进行重启电源。推荐使用 RPS-10 电子开关（美国为 M/HD 模式，欧洲为 M/EC 模式），可从 www.wti.com/rps-10.htm 获得。关于如何在集群中使用电子开关，请参见配置电子开关。 | 是 |
| 空调制解调器线缆 | 2 | 直接串行口线缆用于连接集群系统的串行口和电子开关。这种串行连接可以让每个集群系统都能对另一个系统重启电源。有的电子开关需要不同的线缆。 | 只有使用电子开关时才需要 |
| 加载托架 | 1 | 部分电子开关支持机架安装式配置。 | 只有使用机架安装式电子开关时才需要 |

**表 3. 电子开关硬件**

| 硬件 | 数量 | 说明 | 要求 |
| --- | --- | --- | --- |
| 外部磁盘存储附件 | 1 | 对于生产环境，我们推荐使用单启始端 SCSI 总线或单启始端光纤通道互连来连接集群系统和单控制器（或双控制器）RAID 阵列。如果要使用单启始端总线或互连， RAID 控制器必须有多个主机端口，并能同时访问主机端口的所有逻辑单元。如果某个逻辑单元能在两个控制器之间进行故障切换，则该过程对操作系统必须是透明的。<br><br>对于开发环境来讲，您可以使用多启始端 SCSI 总线或多启始端光纤通道互连来连接集群系统和 JBOD 存储附件；单端口 RAID 阵列、或 RAID 控制器不能访问存储附件端口上的所有逻辑端口。<br><br>您不能在集群中使用基于主机的产品、基于适配器的产品或软件 RAID 产品，因为这些产品通常不能很好地配合多系统接入共享存储器。<br><br>详见配置共享磁盘存储器。 | 否。但为了确保数据完整性，强烈推荐使用。如果有该硬件，我们推荐采用 SCSI 保留支持，将其作为最简单的故障切换并保障数据完整性的解决方案。 |
| 主机总线适配器 | 2 | 如果要连接共享磁盘存储器，您必须在每个集群系统的 PCI 插槽中安装一个并行 SCSI 或一个光纤通道主机总线适配器。<br><br>对于并行 SCSI 来讲，可以采用低电压差分（LVD）主机总线适配器。适配器带有 | 只有采用了外部磁盘存储附件时才需要 |

| 硬件 | 数量 | 说明 | 要求 |
|---|---|---|---|
| | | HD68 或 VHDCI 连接器。如果您需要支持热插拔功能，就必须禁用主机总线适配器的板载终端。推荐使用的并行 SCSI 主机总线适配器有：<br><br>Adaptec 2940U2W、29160、29160LP、39160 和 3950U2<br><br>采用英特尔 L440GX+主板的 Adaptec AIC-7896<br><br>Qlogic QLA1080 和 QLA12160<br><br>Tekram Ultra2 DC-390U2W<br><br>LSI Logic SYM22915<br><br>推荐使用的光纤通道主机总线适配器为 Qlogic QLA2200。<br><br>关于设备特性与配置信息，请参见主机总线适配器的特性与配置要求及自适应主机总线适配器的要求。 | |
| SCSI 线缆 | 2 | 68 针 SCSI 线缆将每台主机总线适配器都连接到存储附件端口上。线缆带有 HD68 或 VHDCI 接头。 | 只用于并行 SCSI |
| 外部 SCSI LVD 主动式终结器 | 2 | 如果要支持热插拔功能，请将一个外部的 LVD 主动式终结器连接到已禁用内部终端的主机总线适配器上。这样，您就可以在不影响总线操作的情况下，将适配器从 | 只有并行 SCSI 保留配置才需要具备外部终端热插拔功能 |

| 硬件 | 数量 | 说明 | 要求 |
| --- | --- | --- | --- |
| | | 终结器上断开。终结器配有 HD68 或 VHDCI 接头。<br><br>我们推荐使用带有 HD68 接头的外部直通式终结器。您可从通过以下途径得到：Technical Cable Concepts， Inc.，350 Lear Avenue ， Costa Mesa ， California， 92626（714-835-1081）。或者请访问 www.techcable.com 网站同他们取得联系。部件说明和编号为：TERM SSM/F LVD/SE Ext Beige，396868-LVD/SE。 | |
| SCSI 终结器 | 2 | 对于使用“ out ”端口并和单启始端 SCSI 总线连在一起的 RAID 存储附件（如：Flashdisk RAID 磁盘阵列）来讲，要终止总线，就得把终结器连接到“ out ”端口上。 | 仅在采用 SCSI 配置和需要终止时使用。 |
| 光纤通道集线器或交换机 | 1 或 2 | 除非您有带两个端口的存储附件，否则必须配备光纤通道集线器或交换机。集群系统中的主机总线适配器可直接连在不同的端口。 | 仅当采用了某些光纤通道配置时才需要。 |
| 光纤通道线缆 | 2 到 6 | 光纤通道线缆用于将主机总线适配器连接到存储附件端口、光纤通道集线器或光纤通道交换机。如果采用了集线器或交换机，则需要额外的线缆把集线器或交换机连接到存储适配器端口。 | 仅当采用了光纤通道配置的时候才需要， |

**表 4. 共享磁盘存储硬件**

| 硬件 | 数量 | 说明 | 要求 |
| --- | --- | --- | --- |
| 网络接口 | 每个网络连接 1 个 | 每个 网络连接需要一个安装在集群系统上的网络接口。有关在集群中使用该驱动的信息请参见 Tulip 网络驱动器的要求。 | 是 |
| 网络交换机或集线器 | 1 | 网络交换机或集线器可使您将多个系统连接到某个网络上。 | 否 |
| 网络线缆 | 每个网络接口使用 1 根网络线缆 | 通用的网络线缆（比如，带有 RJ45 连接器的线缆）可将每个网络接口连接到某台网络交换机或网络集线器上。 | 是 |

**表 5. 网络硬件**

| 硬件 | 数量 | 说明 | 要求 |
| --- | --- | --- | --- |
| 网络接口 | 每个通道 2 个 | 每条以太网心跳通道需要两个集群系统上各安装一个网络接口。 | 否 |
| 网络交叉线缆 | 每个通道使用 1 条 | 网络交叉线缆可将一个集群系统上的某网络接口直接连接到另一个集群系统上的对应网络接口上，从而创建一条以太网心跳通道。 | 只在冗余以太网心跳通道时才需要 |

**表 6. 点对点以太网心跳通道硬件**

| 硬件 | 数量 | 说明 | 要求 |
|---|---|---|---|
| 串行卡 | 每个串行通道使用 2 个 | 每个串行心跳通道在两个集群系统上各需要一个串行口。您可使用多端口串行 PCI 卡来扩展串行口容量。推荐您使用的多端口卡包括以下几种：<br><br>Vision Systems VScom 200H PCI 卡，它提供有两个串行口，详情请访问 www.vscom.de（如欲了解更多信息，请访问 Vscom 多端口串行卡）。<br><br>Cyclades-4YoPCI+卡，它提供有四个串行口，详情请访问 www.cyclades.com。 | 否 |
| 空调制解调器线缆 | 每个通道使用 1 条 | 空调制解调器线缆可将一个集群系统上的某个串行口直接连接到另一个集群系统上的某个对应串行口，从而创建一条串行心跳通道。 | 只对串行心跳通道要求 |

表 7. 点对点串行心跳通道硬件

| 硬件 | 数量 | 说明 | 要求 |
|---|---|---|---|
| 终端服务器 | 1 | 一台终端服务器可帮助您从远程位置管理大量系统。推荐您使用的终端服务器包括以下几种：<br><br>Cyclades 终端服务器，在 www.cyclades.com 上提供。<br><br>NetReach Model CMS-16，由 Western Telematic 公司提供推，请访问 www.wti.com/cms.htm。 | 否 |
| RJ45 到 DB9 交叉线缆 | 2 | RJ45 到 DB9 交叉线缆将每个集群系统上的串行口连接到一个与之对应的 Cyclades 终端服务器上。其它类型的终端服务器可能需要不同的线缆。 | 只对终端服务器要求 |

| 硬件 | 数量 | 说明 | 要求 |
|---|---|---|---|
| 网络线缆 | 1 | 网络线缆可将一台终端服务器连接到与之对应的网络交换机或集线器上。 | 只对终端服务器要求 |
| KVM | 1 | KVM 可使多个系统共享一套键盘、显示器和鼠标。推荐您使用的 KVM 是 Cybex Switchview，详情请访问 www.cybex.com.。将系统连接到交换机的线缆取决于 KVM 的类型。 | 否 |

**表 8. 控制台交换机硬件**

| 硬件 | Quantity | 说明 | 要求 |
|---|---|---|---|
| UPS 系统 | 1 或 2 个 | 不间断的电源（UPS）系统 可提供一种高可用性电源。最理想选择是将用于共享存储附件和两个电子开关的电源线连接到冗余 UPS 系统上。此外，UPS 系统必须能在足够长的时间段内提供电压。<br><br>推荐您使用的 UPS 系统是 APC Smart-UPS 1000VA/670W，由 www.apc.com 网站提供。 | 强烈推荐您使用 |

**表 9. UPS 系统硬件**

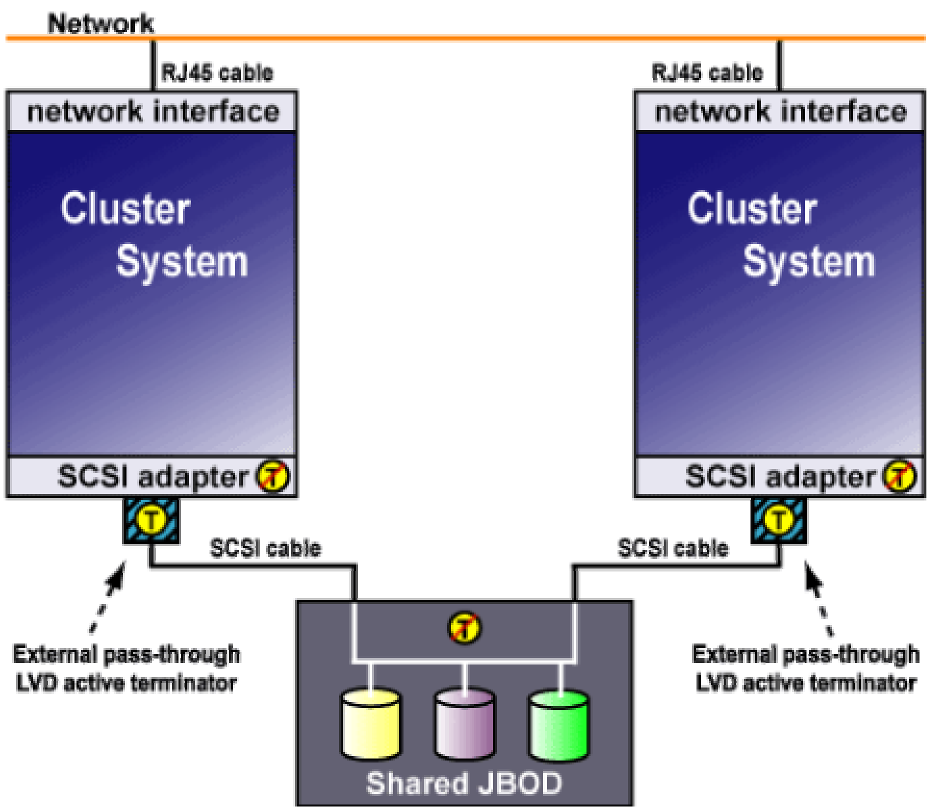
### 2.1.2 最低集群配置举例

下表中介绍的硬件组件可以用来设置最低集群配置（使用一条多启始端 SCSI 总线并支持热插拔功能）。这种配置不能保证所有故障条件下的数据完整性，因为它不包含电子开关。注意，这只是一个配置范例；您可以使用其它硬件设置一种最低配置。

| | |
|---|---|
| 两台服务器 | 每个集群系统包括以下硬件：<br><br>用于客户机访问的网络接口及以太网心跳通道 |
| 两条带有 RJ45 接头的网络线缆 | 网络线缆用于将每个集群系统上的网络接口连接到用于客户机访问和以太网心跳的网络。 |

**表 10. 最低集群硬件配置举例**

下图显示了最低集群硬件配置，包括以上表格中提到硬件和多启始端 SCSI 总线，同样也支持热插拔。被圈住的“T”表示内部（板载）或外部 SCSI 总线终止。“T”上画斜杠表示终止已经被禁止。




带有热插拔的集群硬件最低配置

### 2.1.3 无单点故障配置举例

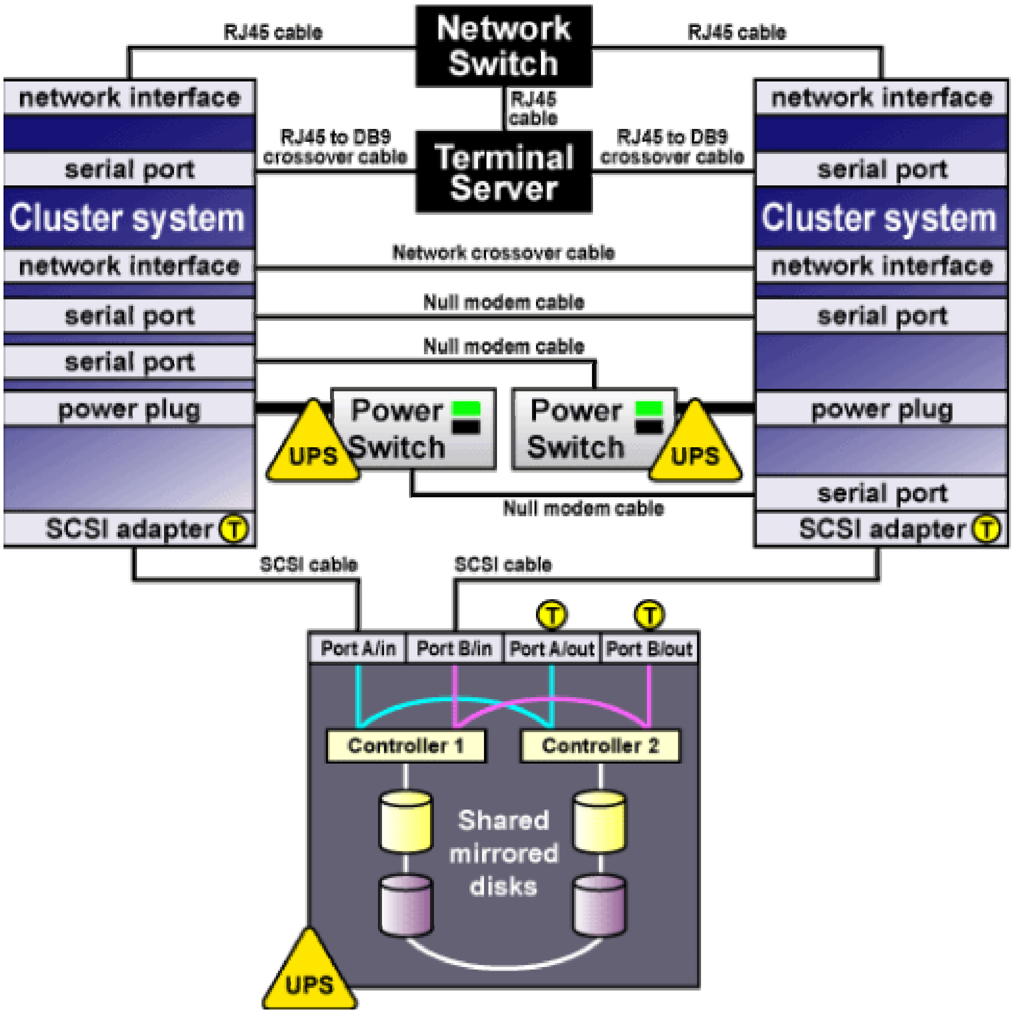
下面表格中的组件，用于设置无单点故障集群配置，它包括两个单启始端SCSI 总线和电子开关，以保证在所有故障条件下数据的完整性。注意，下面只是一个范例，您也可以使用其它硬件设置无单点故障配置。

| | |
|---|---|
| 两个服务器 | 每个集群系统包括以下硬件：<br><br>两个网络接口：<br><br>点对点以太网心跳通道客户机网络接入以及以太网心跳连接<br><br>3 个串行口：<br><br>把点对点串行心跳通道远程电子开关连接到终端服务器。<br><br>一个 Tekram Ultra2 DC-390U2W 适配器（终端已启用），用于共享磁盘存储器连接。 |
| 一个网络交换机 | 使用网络交换机，您可以把多个系统连接到网络中。 |
| 一个 Cyclades 终端服务器 | 使用终端服务器，您就可以在中心位置管理远程系统。 |
| 3 条网络线缆 | 网络线缆可以把位于每个集群系统上的终端服务器和网络接口连接到网络交换机。 |
| 2 条 RJ45 至 DB9 接头的双绞线 | RJ45 接头到 DB9 接头的双绞线可以把每个集群系统的串行口连接到 Cyclades 终端服务器。 |
| 一条网络双绞线 | 网络双绞线可以把两个集群系统的网络接口连接在一起，创建一条点对点以太网心跳通道。 |
| 2 个 RPS-10 电子开关 | 在重启服务之前，电子开关允许每个集群系统给对方集群系统重启电源。每个集群系统的电源线要连接在自己的电子开关上。 |

| | |
|---|---|
| 3 条直接串行口线缆 | 直接串行口线缆可以把每个集群系统的串行口连接到为其他集群系统提供电源的电子开关上。通过这种连接，每个集群系统都能让对方集群系统进行重启电源。<br><br>空调制解调器线缆可将一台集群系统的串行口连接到另一台集群系统的响应串行口上，以此构成一个点到点的串行心跳通道。 |
| 带双控制器的 FlashDisk RAID 磁盘阵列 | 双 RAID 控制器可以防止磁盘和控制器发生故障。 RAID 控制器可以向主机端口所有逻辑单元提供同时接入。 |
| 2 条 HD68 SCSI 线缆 | HD68 线缆可把每个主机总线适配器都连接到 RAID 附件的“ in ”端口，以构成两条单启始端 SCSI 总线。 |
| 2 个终结器 | 连接 RAID 附件的每个“ out ”端口的终结器可以终止双方单启始端 SCSI 总线。 |
| 冗余 UPS 系统 | UPS 系统能提供高可用性电源。电子开关和 RAID 附件的电源线应接在两个 UPS 系统上。 |

**表 11 无单点故障配置举例**

下图为无单点故障硬件配置的一个例子。该配置中包含了上表中所描述的硬件：两条单启始端 SCSI 总线以及在任何故障状态下都能保证数据完整性的电子开关。




无单点故障配置举例

## 2.2 集群系统的设置步骤

硬件组件识别完成以后，依照选择硬件配置中的说明来安装基本的集群系统硬件，并将系统连接到可选控制台交换机和网络交换机（或网络集线器）上。请遵循如下步骤：

1. 在两个集群系统中均安装所需的网络适配器、串行卡和主机总线适配器。参见安装基本系统硬件，了解更多关于该步骤的相关信息。
2. 安装可选控制台交换机，并把它连接到每台集群系统上。参见安装控制台交换机，了解更多关于该步骤的相关信息。

如果您没有采用控制台交换机，那么将每台系统都连接到一个控制台终端上。

3. 安装可选网络交换机或集线器，并使用网线将它连接到集群系统和终端服务器上（如果可用的话）。参见安装网络交换机或集线器，了解更多关于此步骤的相关信息。

   如果您没有使用网络交换机或集线器，就用网线将每台系统和终端服务器（如果可用的话）连接到网络上。

在完成上个步骤后，即可按照安装和配置 Linux 发行版的步骤的描述来安装 Linux 发行版 OS。

### 2.2.1 安装基本系统硬件

集群系统必须提供您应用程序所需的 CPU 处理能力和内存我们建议每套系统至少应配备 450Mhz（或更高）CPU 和 256MB 内存。

此外，集群系统还必须提供您硬件配置所需的 SCSI 适配器、网络接口和串行口。系统应配备一定数量的预装串行口、网络端口以及 PCI 扩展插槽。下表将有助您确定自己的集群系统需要多大的容量：

| 集群硬件组件 | 串行口 | 网络接口插槽 | PCI 插槽 |
| --- | --- | --- | --- |
| 远程电子开关连接（可选） | 一个 | | |
| 到共享磁盘存储器的 SCSI 总线（可选） | | | 每条总线配备一条 |
| 提供客户机访问和以太网心跳通道的网络连接 | | 每条网络连接配备一条 | |
| 点对点以太网心跳通道（可选） | | 每条通道配备一条 | |

| 集群硬件组件 | 串行口 | 网络接口插槽 | PCI 插槽 |
| --- | --- | --- | --- |
| 点对点串行心跳通道（可选） | 每条通道配备一条 | | |
| 终端服务器连接（可选） | 一条 | | |

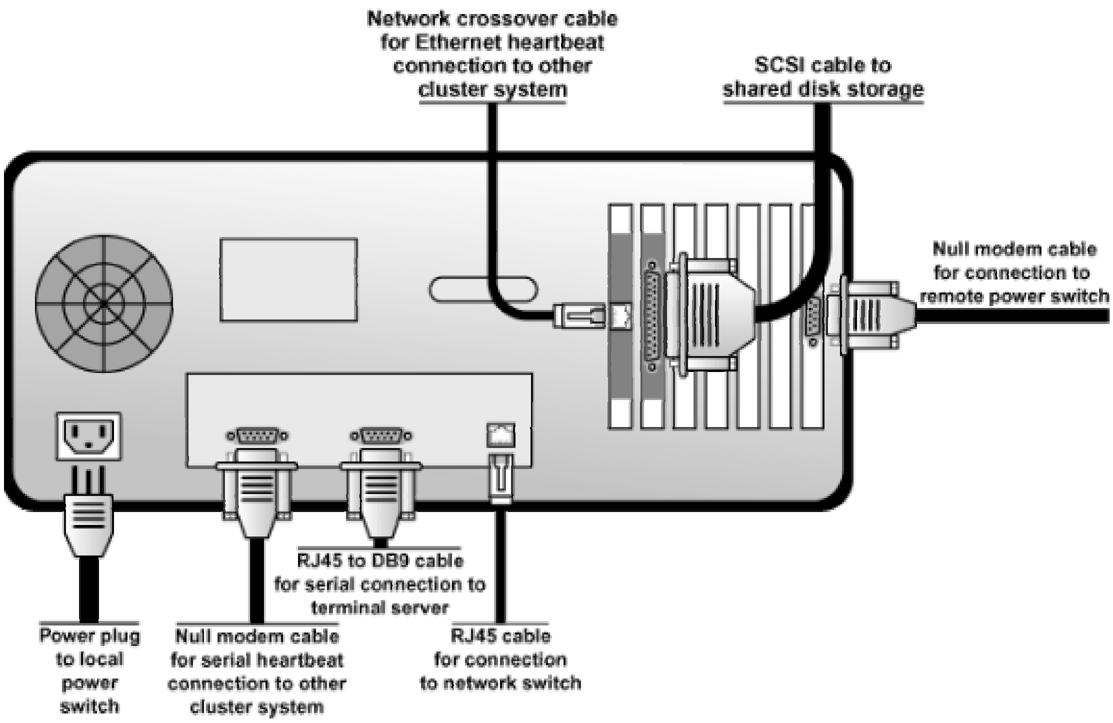
**表 12.集群硬件组件**

大多数系统都至少具备一个串行口，最好选择具有两个以上串行口的系统。如果您的系统具有图形显示功能，您可以将串行控制台端口用于串行心跳通道或者电子开关连接。如果要扩展串行口容量，您可以使用具有多个串行口的串行 PCI 卡。

此外，您必须注意不要将本地系统盘和共享磁盘安装在同一条 SCSI 总线上。要解决这一问题，您可以使用双通道 SCSI 适配器（如 Adaptec 3950-系列串行卡），将内部设备安装在其中一条通道上，将共享磁盘安装在另一条通道上，或者您也可使用多个 SCSI 接口卡。

具体安装细节请请参见厂商提供的系统文档。关于如何在集群中使用主机总线适配器、多串行口卡以及 Tulip 网络驱动程序，请请参见补充硬件信息

下图显示了集群系统的样板机房以及标准集群配置的外部电缆连接。




标准集群系统的外部电缆连接

## 2.2.2 安装控制台交换机

虽然集群操作不是必须都要使用控制台交换机，但是使用它能够方便您进行系统管理，让您不必为每个集群系统都单独配备显示器、鼠标和键盘。控制台交换机分为几种不同类型。

例如，您可以通过终端服务器连接到串行控制台，远程管理多个系统。为了降低成本，您还可以使用 KVM（键盘、显示器、鼠标）交换机，在多个系统间共用一套键盘、显示器和鼠标。KVM 交换机非常适合于通过图形用户界面（GUI）执行系统管理任务的配置情形。

如本手册中未另外提供专门的集群安装详细指导，请依照厂商提供的文档安装控制台交换机。

安装完控制台交换机后，再把它连接到各个集群系统上。

### 2.2.3 安装网络交换机或集线器

虽然集群操作并不是必须使用网络交换机或者集线器,但使用它能够方便您的集群和客户机系统网络操作。

请依照厂商提供的文档来安装网络交换机或者集线器。

装完网络交换机集线器后，请使用通常的网线把它连接到各个集群系统上。如果您使用了终端服务器，请使用网线把它连接到网络交换机或者网络集线器上。

## 2.3 安装和配置 Linux 发行版 OS

完成基本硬件的安装之后，请分别在两个集群系统上安装 Linux 发行版，并确保它们能够识别所连接的设备。请遵循如下步骤：

1. 依照安装 Linux 发行版中所描述的内核要求和指导在两个集群系统上安装 Linux 发行版。
2. 重新启动集群系统。
3. 如果您使用终端服务器，请配置 Linux 向控制台端口发送控制台信息。

   如果您使用的是 Cyclades 终端服务器，请请参见配置 TurboLinux，以向控制台端口发送控制台信息来获得有关此项操作的更多信息。

4. 编辑每个集群系统中的/etc/hosts 文件，并把集群中用到的 IP 地址包含在内。请参见编辑/etc/hosts 文件来获得有关此项操作的更多信息。
5. 减小内核交替启动超时限制，降低系统的启动时间。请参见降低内核交替启动超时限制来获得有关此项操作的更多信息。
6. 确保登录程序没有使用用于心跳通道或远程电子开关连接（如果有的话）的串行口。为此，您可以编辑/etc/inittab 文件，使用数字标记（#）来分别注释对应于串行通道和远程电子开关串行口的启动行。然后，调用 init q 命令。

7. 确认两个系统都检测到了所有已安装的硬件：
   - 使用 dmesg 命令显示控制台启动信息。请参见显示控制台启动信息，了解有关此项操作的更多信息。
   - 使用 cat /proc/devices 命令来显示内核中配置的设备。请参见确定内核中已配置的设备，了解有关此项操作的更多信息。
8. 使用 ping 命令将测试数据包从一个系统发送到另一个系统，以确定集群系统可以和所有的网络接口进行通讯。

### 2.3.1 安装 GTES10 操作系统

调度节点需要安装 GTES10 操作系统，真实服务器需要支持所选择类型的要求。比如，如果用户选择 TUN 方式，那么 realserver 需要能够加载 ipip 模块。在安装 Linux 发行版之前，您先要获得集群系统和端对端以太网心跳接口的 IP 地址。端对端以太网接口的 IP 地址可以是专有 IP 地址，如：10.0.0.x。

在安装 Turbolinux Server 时，请注意以下事项：

- 不要将系统文件系统（例如：/、/usr、/tmp、和/var）放在共享磁盘上。
- 将/tmp 和/var 放在不同的文件系统中。

启动顺序

您的引导盘必须是系统中第一个识别的磁盘，IDE 磁盘通常先于 SCSI 磁盘被识别为/dev/hda。如果您的引导磁盘是 IDE 型而您的共享磁盘是 SCSI 型，那么您的系统就不存在启动顺序的问题，您可以跳过这一段。

如果您的引导磁盘和您的共享磁盘都是 SCSI 设备，那么您可能需要修改 SCSI 控制器的启动顺序才能确保您的系统盘比共享磁盘先被识别。引导磁盘必须始终被识别为/dev/sda。要改变 SCSI 控制器的启动顺序，您可以：改变主板 BIOS 中关于 PCI 扫描顺序的设置，将共享存储控制器的 BIOS 设置改为不将其作为引导设备，或者改变插件板在您主板 PCI 插槽中的位置次序。

### 2.3.2 编辑/etc/hosts 文件

/etc/hosts 文件中包含 IP 地址到主机名的转换表。在每个集群系统中，该文件都必需包含以下内容项：

- 两个集群系统使用的 IP 地址及其对应的主机名称。
- 以太网心跳连接（可以是专有 IP 地址）使用的 IP 地址及其对应的主机名称。

下面是一个/etc/hosts 文件举例：

| | | |
|---|---|---|
| 127.0.0.1 | localhost.localdomain | localhost |
| 193.186.1.81 | cluster2.linux.com | cluster2 |
| 10.0.0.1 | ecluster2.linux.com | ecluster2 |
| 193.186.1.82 | cluster3.linux.com | cluster3 |
| 10.0.0.2 | ecluster3.linux.com | ecluster3 |

**表 13. /etc/hosts**

您可以使用 DNS 替代/etc/hosts 文件来解析您网络上的主机名称。

上面的举例显示了两个集群系统的 IP 地址和主机名称（cluster2 和 cluster3）以及每个集群系统（ecluster2 和 ecluster3）中点对点心跳连接所用以太网接口的 IP 地址和主机名。

下面是在一个集群系统中使用 ifconfig 命令的部分输出结果：

| | |
|---|---|
| eth0 | Link encap:Ethernet HWaddr 00:00:BC:11:76:93<br>inet addr:192.186.1.81 Bcast:192.186.1.245 Mask:255.255.255.0<br>UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1<br>RX packets:65508254 errors:225 dropped:0 overruns:2 frame:0 |

| | |
|---|---|
| | TX packets:40364135 errors:0 dropped:0 overruns:0 carrier:0<br>collisions:0 txqueuelen:100<br>Interrupt:19 Base address:0xfce0 |
| eth1 | Link encap:Ethernet HWaddr 00:00:BC:11:76:92<br>inet addr:10.0.0.1 Bcast:10.0.0.245 Mask:255.255.255.0<br>UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1<br>RX packets:0 errors:0 dropped:0 overruns:0 frame:0<br>TX packets:0 errors:0 dropped:0 overruns:0 carrier:0<br>collisions:0 txqueuelen:100<br>Interrupt:18 Base address:0xfcc0 |

**表 14. #ifconfig**

上例中显示了一个集群系统中的两个网络接口：eth0（集群系统的网络接口）和 eth1（点对点心跳连接网络接口）。

编辑 Root $PATH 变量：我们推荐您在新安装的或者现有的 Turbolinux Server 中，编辑 root $PATH 变量来将/opt/cluster/bin 添加到 bashrc $PATH 中。

## 2.3.3 显示控制台启动信息

使用 dmesg 命令来显示控制台启动信息。请参见 dmesg（8）manpage 了解更多信息。

下面的 dmesg 命令输出显示系统在启动过程中识别出 1 个串行扩展卡：

May 22 14：02：10 storage3 kernel：Cyclades driver 2.3.2.5 2000/01/19 14：35：33

May 22 14：02：10 storage3 kernel：built May 8 2000 12：40：12

May 22 14：02：10 storage3 kernel：Cyclom-Y/PCI #1：0xd0002000-0xd0005fff, IRQ9,

4 channels starting from port 0.

下面的 dmesg 命令输出显示在系统上检测到 2 条外部 SCSI 总线和 9 个磁盘：

May 22 14：02：10 storage3 kernel：scsi0：Adaptec AHA274x/284x/294x（EISA/VLB/PCI-Fast SCSI）5.1.28/3.2.4

May 22 14：02：10 storage3 kernel：

May 22 14：02：10 storage3 kernel：scsi1：Adaptec AHA274x/284x/294x（EISA/VLB/PCI-Fast SCSI）5.1.28/3.2.4

May 22 14：02：10 storage3 kernel：

May 22 14：02：10 storage3 kernel：scsi：2 hosts.

May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST39236LW Rev：0004

May 22 14：02：11 storage3 kernel：Detected scsi disk sda at scsi0, channel 0, id 0, lun 0

May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST318203LC Rev：0001

May 22 14：02：11 storage3 kernel：Detected scsi disk sdb at scsi1, channel 0, id 0, lun 0

May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST318203LC Rev：0001

May 22 14：02：11 storage3 kernel：Detected scsi disk sdc at scsi1, channel 0, id 1, lun 0

May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST318203LC Rev：0001

May 22 14：02：11 storage3 kernel：Detected scsi disk sdd at scsi1, channel 0, id 2, lun 0

May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST318203LC Rev：0001

May 22 14：02：11 storage3 kernel：Detected scsi disk sde at scsi1, channel 0, id 3, lun 0

May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST318203LC Rev：0001

May 22 14：02：11 storage3 kernel：Detected scsi disk sdf at scsi1, channel 0, id 8, lun 0

```
May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST318203LC Rev： 0001
May 22 14：02：11 storage3 kernel：Detected scsi disk sdg at scsi1, channel 0, id 9, lun 0
May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST318203LC Rev： 0001
May 22 14：02：11 storage3 kernel：Detected scsi disk sdh at scsi1, channel 0, id 10, lun 0
May 22 14：02：11 storage3 kernel：Vendor：SEAGATE Model：ST318203LC Rev： 0001
May 22 14：02：11 storage3 kernel：Detected scsi disk sdi at scsi1, channel 0, id 11, lun 0
May 22 14：02：11 storage3 kernel：Vendor：Dell Model：8 BAY U2W CU Rev： 0205
May 22 14：02：11 storage3 kernel：Type：Processor ANSI SCSI revision： 03
May 22 14：02：11 storage3 kernel：scsi1：channel 0 target 15 lun 1 request sense failed, performing reset.
May 22 14：02：11 storage3 kernel：SCSI bus is being reset for host 1 channel 0.
May 22 14：02：11 storage3 kernel：scsi：detected 9 SCSI disks total.
```

下面的 dmesg 命令输出显示在系统上检测到 1 个 Quad 以太网卡：

```
May 22 14：02：11 storage3 kernel：3c59x.c：v0.99H 11/17/98 Donald Becker http：//cesdis.gsfc.nasa.gov/linux/drivers/vortex.html
May 22 14：02：11 storage3 kernel：tulip.c：v0.91g-ppc 7/16/99 becker@cesdis.gsfc.nasa.gov
May 22 14：02：11 storage3 kernel：eth0：Digital DS21140 Tulip rev 34 at 0x9800, 00：00：BC：11：76：93, IRQ 5.
May 22 14：02：12 storage3 kernel：eth1：Digital DS21140 Tulip rev 34 at 0x9400, 00：00：BC：11：76：92, IRQ 9.
May 22 14：02：12 storage3 kernel：eth2：Digital DS21140 Tulip rev 34 at 0x9000, 00：00：BC：11：76：91, IRQ 11.
May 22 14：02：12 storage3 kernel：eth3：Digital DS21140 Tulip rev 34 at 0x8800, 00：
```

```
00：BC：11：76：90, IRQ 10.
```

### 2.3.4 确定内核中已配置的设备

在每个集群系统上使用 cat /proc/devices 命令,确认所安装的设备（包括串行口和网络接口）是否已在内核中配置完成。例如：

```
# cat /proc/devices
Character devices：
1 mem
2 pty
3 ttyp
4 ttyS [1]
5 cua
7 vcs
10 misc
19 ttyC [2]
20 cub
128 ptm
136 pts
162 raw [3]
Block devices：
2 fd
3 ide0
8 sd [4]
65 sd
#
```

上例中显示了如下内容：

[1] Onboard serial ports（ttyS）

[2] Serial expansion card（ttyC）

[3] Raw devices（raw）

[4] SCSI devices（sd）

## 2.4 安装和连接集群硬件的步骤

在 Linux 发行版安装完毕之后，您就可以安装集群硬件组件了，然后确认集群系统是否识别出了所有已连接的设备。注：具体安装步骤取决于配置类型。请参见选择硬件配置，了解关于集群配置的更多信息。

请按照下列步骤来安装集群硬件：

1. 关闭集群系统，并断开电源。
2. 设置点对点以太网和串行心跳通道（如果需要的话）。请参见配置心跳通道，了解有关此项操作的更多信息。
3. 如果您使用的是电子开关，在设备安装完成后，将各个集群系统连接到电子开关上。注：在集群中使用电子开关，您也许要设定循环地址或拨动开关。请参见配置电子开关，了解有关此项操作的更多信息。

   此外，我们推荐将各电子开关（如果没有使用电子开关的话，此处应为电源线）连接到不同的 UPS 系统上。请参见配置 UPS 系统，了解有关使用 UPS 系统的更多信息。

4. 依照厂商的说明书来安装共享磁盘，并将集群系统连接到外部存储设备上，注意确保满足多启始端或者单启始端 SCSI 总线的配置要求。请参见配置共享磁盘，了解有关此项操作的更多信息。

   此外，我们建议您将存储设备连接到冗余 UPS 系统上。请参见配置 UPS 系统，了解关于使用可选 UPS 系统的更多信息。

5. 接通硬件电源，启动每个集群系统。在启动过程中，进入 BIOS 工具，依照如下方法修改系统设置：
    - 为 SCSI 总线上的每条主机总线适配器指定一个唯一的 SCSI 标识号。请参见 **SCSI 标识号**，了解有关此项操作的更多信息。
    - 根据您的存储器配置，启用或停用各主机总线适配器的板载终端。请参见**配置共享磁盘和 SCSI 总线终端**，了解有关此项操作的更多信息。
    - 如果您的主机总线适配器能够正确进行总线复位并且内核为 2.2.18 或者更高版本，那么您就可以把总线复位保持在启用状态。2.2.18Adaptec 驱动程序不支持总线复位和当前 TurboLinux 发行版包括 2.2.18 版本或者更高版本的内核。
    - 启用集群系统的加电自动启动特性。

如果您的共享存储器使用的是 Adaptec 主机总线适配器，请参见 **Adaptec** 主机总线适配器的要求，了解配置信息。

1. 退出 BIOS 工具，继续启动每个系统。检查启动信息，确认 Linux 内核是否已经配置完成，并能够识别整套共享磁盘。您也可以使用 dmesg 命令来显示控制台启动信息。请参见显示控制台启动信息，了解更多关于本命令的相关信息。
2. 使用 ping 命令向各个网络接口发送数据包，确认集群系统是否可以与每个点对点以太网心跳连接进行通信。

### 2.4.1 配置心跳通道

集群系统使用心跳通道来互相通信，并确定集群系统的状态。例如，如果一个集群系统停止通过心跳通道来更新其时间戳，另一个集群系统就会检查心跳通道的状态以决定是否需要进行故障切换。

每个集群必须包括至少一条心跳通道，您可以将一条以太网连接既用于客户机访问，又用于心跳通道。然而，我们建议您还是另外安装别的心跳通道以提高

系统的可用性。除了一条或多条串行心跳通道之外，您还可以安装冗余以太网心跳通道。

例如，如果您拥有一条以太网通道和一条串行心跳通道，那么即使以太网通道电缆断开，集群系统仍然可以通过串行心跳通道来检查状态。

要建立冗余以太网心跳通道，您可以使用一个双绞线来将一个集群系统上的网络接口连接到另一个集群系统的网络接口上。

要建立串行心跳通道，使用一条串行空调制解调器电缆将一个集群系统的串行口连接到另一个集群系统的串行口上即可。确保连接到集群系统上对应的串行口，不要连接到将用于远程电子开关连接的串行口上。

### 2.4.2 配置电子开关

电子开关使集群系统在故障切换过程中可先对对方集群系统重启电源，随后再重启其服务。远程系统停用能力可确保任何故障情况下数据的完整性。我们建议生产环境下应在集群配置中使用电子开关，无电子开关配置仅供用于开发环境。

在使用电子开关的集群配置中，每个集群系统的电源线均连接到各自的电子开关上。此外，两个集群系统通常还使用串行口连接的方式远程连接到对方的电子开关上。一旦发生故障切换，集群系统可使用该连接在重启服务之前，先对另一个集群系统重启电源。

电子开关可以防止未响应（“挂起”）系统在故障切换后恢复响应（“解除挂起”）时可能会产生的数据损坏，并对正从另一个集群系统接收 I/O 数据的磁盘做 I/O 操作。

此外，如果一个集群系统中的 quorum daemon 发生故障，该系统将不能够再监视 quorum partition。如果您在集群中没有使用电子开关，这种故障会造成服务运行在不止一个集群系统中，最终导致数据损坏。

我们推荐您使用电子开关或者集群中的SCSI保留功能来确保故障切换后数据的完整性。SCSI 保留应优先考虑，应为它无需额外的硬件成本和安装工作。请查阅有关 SCSI 保留的共享存储器配置部分。如果您决定使用电子开关，则必须指定-p 选项来启用。

此时，如果集群系统正在进行页面交换或者系统工作负载较高，它会“ 挂起 ”几秒钟。这种情况下，不会进行故障切换，因为另一个集群系统并不能断定“ 挂起 ” 系统是否关闭。

集群系统可能会因为硬件故障或者内核错误而挂起。如果发生这种情况，另一个集群系统会察觉到 “ 挂起 ” 的系统不再更新其 quorum partitions 上的时间戳，也不再响应心跳通道上的 ping 命令了。

如果集群系统断定挂起的系统已关闭，在采用电子开关的情形中，该集群系统会在重启其服务之前，先对挂起系统进行重启电源。这就使得挂起系统能够得以 “ 干干净净 ” 地重启，防止发生 I/O 指定与服务数据损坏错误。

如果在集群中没有使用电子开关，那么当集群系统断定某个挂起系统已关闭后，将在 quorum partitions 上将故障系统的状态设定为 DOWN，然后重新启动挂起系统的服务。如果挂起系统解除挂起，它会注意到自己的状态为 DOWN，因而启动系统重启。这样就最大限度缩短了两个集群系统向同一磁盘发送数据的时间间隔，但是这种方式并不能提供电子开关能够保障的数据完整性。如果挂起系统一直不恢复响应，那么您将不得不人工重启系统。

如果您使用电子开关，请参照厂商的说明书来安装硬件。此外，要在集群中使用电子开关，您可能还需要执行一些特殊的操作。请参见设置 RPS-10 电子开关，了解更多关于如何在集群中使用 RPS-10 电子开关的信息。注：如果本手册

中也提供有同样的操作信息，以本手册为准。在安装集群软件时要记得在 member_config 命令中指定-p 选项。

安装完电子开关之后，执行下列步骤，将它们连接到集群系统中：

1. 将每个集群系统的电源线连接到电子开关上。
2. 将每个集群系统的一个串行口连接到向另外一个集群系统供电的电子开关的串行口上。串行口连接所用的电缆与电子开关的类型有关。例如，如果您使用的是 RPS-10 型电子开关，则选择空调制解调器电缆。
3. 将每个电子开关的电源线接到电源上。我们建议您将各个电子开关连接到不同的 UPS 系统上。请参见配置 UPS 系统，了解更多信息。

当您安装完集群软件后，在启动集群前，请先对电子开关进行测试以确保每个集群系统都能对另一个系统进行电源重启。请参见测试电子开关，了解更多信息。

### 2.4.3 配置 UPS 系统

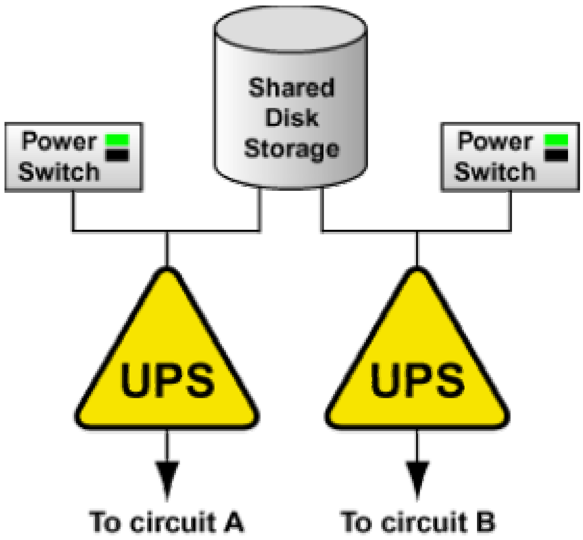
不间断电源（UPS）系统可在发生停电事故时继续供电，以避免系统停机。虽然使用 UPS 系统对集群操作来说不是必需的，但我们仍然建议您这样做。为了获得最高的可用性，请将电子开关（或者电源线－没有使用电子开关时）和磁盘存储子系统都连接到冗余 UPS 系统上。此外，每个 UPS 系统必须连接它自身的电源电路中。

同时，您还要确保每个 UPS 系统都能够为其连接设备提供充足的电力。一旦发生停电，UPS 系统必须保持适当长时间的连续供电。

冗余 UPS 系统可以提供高度可用的电源。一旦发生停电，集群设备的功率负载会在多个 UPS 系统之间进行分配。这样，即使一个 UPS 系统发生故障，集群应用也依然可用。

如果您的磁盘存储子系统备有两套电源，各自有着独立的电源线，那么您需要设置两台 UPS 系统，每台均连接一个电子开关（或一个集群系统的电源线，如果没有采用电子开关的话）和存储子系统的一条电源线。

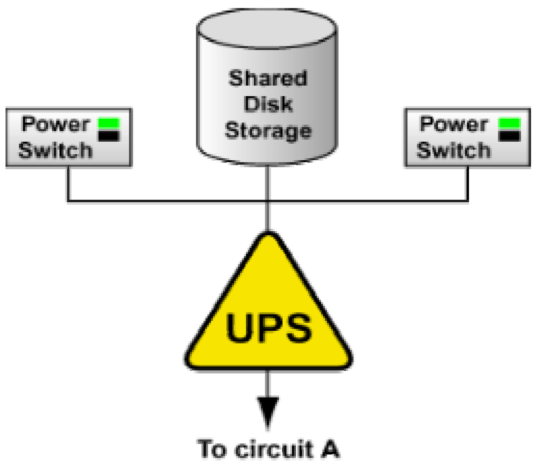
下图显示的是冗余 UPS 系统配置。




冗余 UPS 系统配置

您也可以将电子开关（或两个集群系统的电源线）和磁盘存储子系统连接到同一个 UPS 系统上。这是最为经济高效的一致配置，能够有效防范各种电源故障。但是一旦发生停电，单个 UPS 系统就会成为可能的单点故障点。此外，一台 UPS 系统也许不能为所有连接设备提供足够长时间的供电。

下图显示的是单个 UPS 系统配置。



单个 UPS 系统配置

许多 UPS 系统产品，包括 Linux 应用程序，可以通过串行口连接来监视 UPS 系统的运行状态。如果电池电量过低，监视软件就会启动彻底的系统关闭。一旦发生这种情况，集群软件会正确关闭，因为它受 System V 运行级脚本（如 /etc/rc.d/init.d/cluster）的控制。

请参见厂商**提供的 UPS 文档**，了解更多安装信息。

### 2.4.4 配置共享磁盘存储器

在集群中，共享磁盘存储器用于保存服务数据。因为该存储器必须保证对两个集群系统随时可用，因而不能安装在依赖某个系统才可用的磁盘上。请参见厂商提供的文档，了解更多产品与安装信息。

当您在集群中设置共享磁盘存储器时应注意以下事项：

- 硬件 RAID 与 JBOD

  JBOD（“简单磁盘组〔just a bunch of disks〕”）存储是一种低成本存储解决方案。但是数据的可用性不是很高。如果 JBOD 中的某一磁盘发生故障，所有使用该磁盘的集群服务都会变得不可用。因此，只有在开发环境中才会使用 JBOD。

  基于控制器的硬件 RAID 比起 JBOD 存储来讲，价格比较高，但是它可以避免磁盘故障造成的影响。此外，双控制器 RAID 阵列还可以避免因控制器故障而造成的影响。我们强烈建议您使用 RAID 1（镜像），确保为您的服务数据和 quorum partitions 提供最高的可用性。

  另外，您也可以采用奇偶校验 RAID 来提高可用性。quorum partitions 不能采用 RAID 0（带区）。我们推荐在产品环境下使用 RAID 来实现高可用性。

注：您不能在集群中使用基于主机的 RAID、基于适配器的 RAID 或者软件 RAID，因为这些产品通常不能正确地处理到共享存储器的多系统访问。

- 多启始端 SCSI 总线或光纤通道互连与单启始端总线或互连

多启始端 SCSI 总线或光纤通道互连连接多个集群系统。带有一个主机端口和多个并行 SCSI 磁盘的 RAID 控制器必须使用多启始端总线或互连来将两个主机总线适配器连接到存储器上。这种配置不支持主机隔离。因此，只有开发环境才采用多启始端总线。

单启始端 SCSI 总线者光纤通道互连仅连接一个集群系统，提供主机隔离功能，与多启始端总线相比，有着更好的性能。单启始端总线或互连可确保每个集群系统都受到保护，免受因另一个集群系统的工作负载、初始化、或者维修所带来的影响。

如果您的 RAID 阵列具有多个主机端口，并能提供从这些端口到所有共享逻辑单元的并发访问，那么您可以设置两条单启始端总线或互连来将每个集群系统都连接到 RAID 阵列上。如果一个逻辑单元可以从一个控制器故障切换到另一个控制器，该过程对于操作系统来讲必须保持透明。我们推荐在生产环境中使用单启始端总线或互连。

- 热插拔

在某些情况下，您还可以设置支持热插拔的共享存储器配置，它使您可以在不影响总线运行的情况下从多启始端 SCSI 总线或者多启始端或光纤通道互连上断开设备。这样一来，您就能够在保持使用总线或互连的服务之可用性的同时，轻松执行设备维护工作了。

例如，通过采用外部终端器来以主机总线适配器的方式代替板载终端作为 SCSI 总线的终端，您可以从适配器上将 SCSI 电缆和终端器从适配器上断开，同时保持总线的终止状态。

但是，如果您采用的是光纤通道总线或者交换机，热插拔功能就没有太大必要了，因为网络集线器或交换机允许在某个设备断开后仍然保持互连的可用性。另外，如果您使用的是单启始端 SCSI 总线或者光纤通道互连，也没有必要使用热插拔功能，因为该专用总线在您断开设备时不需要保持可用性。

- SCSI 保留

为了确保故障切换后的数据完整性，我们建议为每个电子开关或者共享存储器配备 SCSI 保留支持能力。GreatTurbo HA 10 在没有电子开关和 SCSI 保留的情况下也可以工作，但是故障切换能力会降低为‘ 软件重启动 ”（请参见下面部分）。因此我们强烈建议在每个 GreatTurbo HA 10 集群中都使用电子开关或者 SCSI 保留，或者两者均使用。默认情况下，在您运行 member_config 并为共享存储磁盘指定 SG 设备时就会用到 SCSI 保留。

SCSI 保留是一种比较简单经济的数据完整性保护方法，因为它无需额外的硬件支持，但同时有一些 “ 折扣 ” 需要注意：1} SCSI 保留仅提供 I/O 隔离或者数据保护，也就是说，它会阻断故障节点到共享存储器的访问，却不复位故障节点。2} SCSI 保留要求每个集群系统内核都包含 “ SCSI 保留 ” 补丁，以避免集群系统在总线复位时错误地清除掉了 SCSI 保留。最新的 Turbolinux 2.2.18 内核包含有 “ SCSI 保留 ” 补丁。3} 共享存储器必须支持 SCSI 保留。GreatTurbo HA 10 中有一个工具可以检查您的共享存储器是否支持 SCSI 保留。下面显示了您可以如何对共享存储器进行测试。

下面是支持 SCSI 保留的共享存储器设备的正常测试结果。

```
================== Try reservation on one node ===================
[root@server1 GreatTurbo HA]# /opt/cluster/bin/sg_switch -d /dev/sg0 -s
/dev/sg0 reserved.
```

```
================= On the other node, test for reservation conflict =======
[root@server2 GreatTurbo HA]# /opt/cluster/bin/sg_switch -d /dev/sg0 -c
Reservation conflict on /dev/sg0.
=================== On the other node, no reservation conflict =====
[root@server2 GreatTurbo HA]# /opt/cluster/bin/sg_switch -d /dev/sg0 -c
No conflict on /dev/sg0.
```

在使用 SCSI 保留时，必须在总线适配器的 BIOS 中启用总线复位特性。在故障节点复位之后，必须清除 SCSI 保留。SCSI 保留的目的就是防止故障系统在未知状态下向共享存储器写入数据。Svcmgr 在尝试向共享存储器写入之前将使用锁定共享存储器来安全地恢复服务。

有时各节点中 sg 设备名称各不相同。您需要在各节点中设定一个指向该设备的符号链接。例如：

```
[root@server1 RPMS]# sg_map
/dev/sg0    /dev/sda

[root@server2 RPMS]# sg_map
/dev/sg1    /dev/sda

[root@server1 RPMS]#ln -s /dev/sg0 /dev/sg8

[root@server2 RPMS]#ln -s /dev/sg1 /dev/sg8
```

- 软件重启动（Software Reboot ）

如果在集群中没有使用电子开关而且共享存储器也不支持SCSI保留，那么您可以采用一种反向故障切换数据完整性保障机制，我们称之为软件重启动。当 power daemon 受命解决一个节点时，它会向故障

节点的 power daemon 发送一条信息触发重启。故障节点在确认之后将立即重启。收到确认信息之后，正常节点会将共享存储器和服务切换到恢复的故障节点。

如果软件重启动失败，正常节点会继续尝试解决故障节点，共享存储和服务也不会切换到故障节点上。如果您在使用 member_config 命令时没有为共享存储器指定 SG 设备，软件重启动模式将默认处于启用状态。为了减少引起故障切换的故障次数，我们强烈建议您使用电子开关或者 SCSI 保留。

注意，为了保证集群能够正常工作，请您务必仔细按照配置说明来配置多启始端和单启始端总线以及热插拔功能。

您必须遵守以下共享存储器要求：

- 各集群系统中每个共享存储设备的 Linux 设备名称必须相同。例如，一个集群系统中名为 /dev/sdc 的设备在另一个集群系统中也必需命名为 /dev/sdc。您可以在两个集群系统中使用相同的硬件以确保设备名称相同。
- 磁盘分区只能供一种集群服务使用。
- 不要在集群系统的本地/etc/fstab 文件夹中包含任何用于集群服务的文件系统 ，因为集群软件必须控制服务文件系统的加载和卸载。
- 为了实现最佳的性能，应在创建共享文件系统时使用 4 KB 的块大小。注意，某些 mkfs 文件系统创建程序默认的是 1 KB 字区大小，这可能会导致较长的 fsck 时间。我们建议您采用像 Reiser 文件系统（reiserfs）或 EXT3 文件系统（ext3）那样的日志文件系统，以消除 fsck 时间，最终缩短故障切换时间。最新版本的 Turbolinux Server 同时支持 reiserfs 和 ext3 两种文件系统，可与 GreatTurbo HA 10 集群故障切换共用。

如果可行的话，您必须遵守下列并行 SCSI 要求：

- SCSI 总线必须在两端终止，且必须遵守长度和热插拔限制。

- SCSI 总线上的设备（磁盘、主机总线适配器和 RAID 控制器）必须具有唯一的 SCSI 标识号。
- 如果使用了 SCSI 保留，则必须启用 SCSI 总线复位功能。

请参见 **SCSI 总线配置要求**，了解更多信息。

另外，我们强烈建议您将存储器连接到冗余 UPS 系统上，以获得高可用性的电源。请参见**配置 UPS 系统**，了解更多信息。

请参见**设置多启始端 SCSI 总线**、**设置单启始端 SCSI 总线**和**设置单启始端光纤通道互连**，了解更多有关配置共享存储器的信息。

安装完共享存储器硬件之后，您可以对磁盘进行分区，之后在分区上创建文件系统或者裸设备。请参见**磁盘分区**、**创建裸设备**、**创建文件系统**，来获得更多信息。

#### 2.4.4.1 设置多启始端 SCSI 总线

多启始端 SCSI 总线可以连接多个集群系统。如果您采用的是 JBOD 存储器，则必须使用多启始端 SCSI 总线来将集群系统连接到集群存储器的共享磁盘上。如果您的 RAID 控制器不能提供从存储器主机端口到所有共享逻辑单元的访问或者它仅有一个主机端口，那么您也必需使用多启始端 SCSI 总线。

多启始端 SCSI 总线不提供主机隔离功能。因此，它只能用于开发环境中。

多启始端总线必须满足 SCSI 总线配置要求中所描述的各项要求。此外，请参见主机总线适配器特性与配置要求，了解有关终止主机总线适配器和配置多启始端总线（具备或者不具备热插拔功能）的更多信息。

一般来说，要设置一条总线两端各有一个集群系统的多启始端 SCSI 总线，您必须遵照以下步骤：

- 启用各主机总线适配器的板载终端。
- 如果可以的话，停用存储器的终端。

- 使用适当的 68-针 SCSI 电缆将每个主机总线适配器连接到存储器。

要设置主机总线适配器终端，您必须在系统启动过程中先进入系统配置工具界面。要设置 RAID 控制器或者存储器终端，请请参见厂商提供的文档。

下图显示了一条不带热插拔支持能力的多启始端 SCSI 总线。

多启始端 SCSI 总线配置

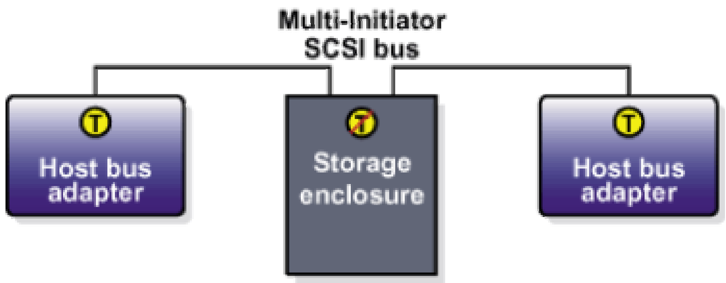
如果主机总线适配器的板载终端可以停用，您可将其配置为支持热插拔。这使您可以将适配器从多启始端 SCSI 总线上断开而不会对总线终端造成影响，从而可以在保持总线正常运行的同时来进行维护操作。

要配置主机总线适配器的热插拔能力，您必须遵照以下步骤来进行：

- 停用主机总线适配器的板载终端。
- 将一个外部通道式（pass-through）LVD 主动式终端器连接到主机总线适配器的接头上。

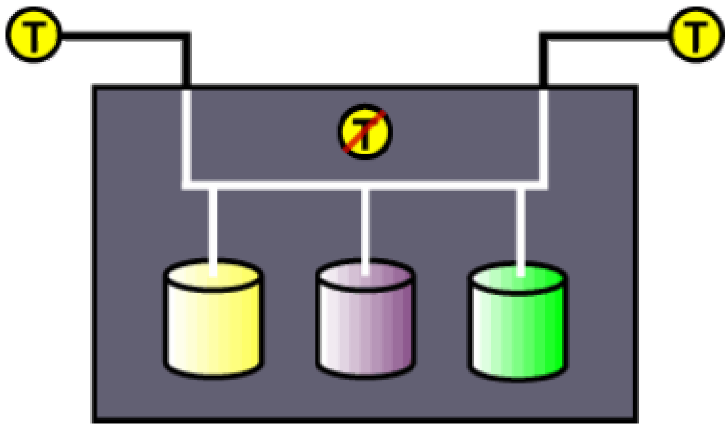
然后使用适当的 68-针 SCSI 电缆将 LVD 终端器连接到（未终止的）存储器上。

下图显示了一条多启始端SCSI总线，其两个主机总线适配器均支持热插拔。



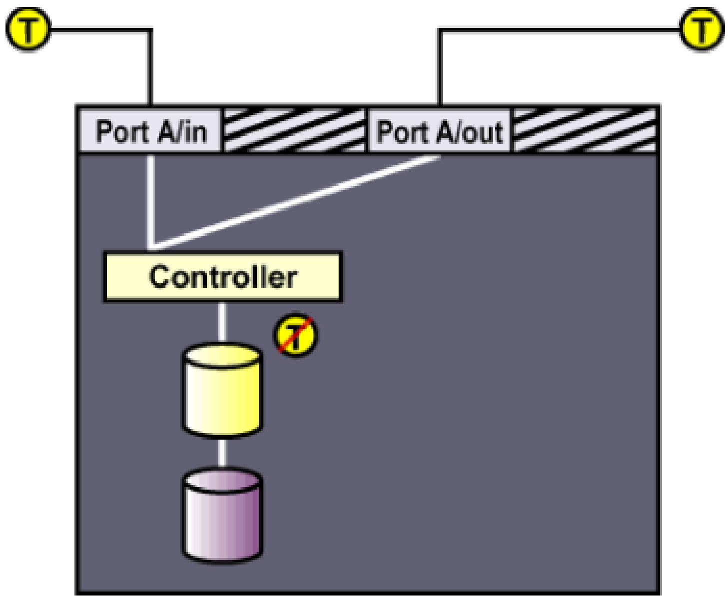

支持热插拔的多启始端 SCSI 总线配置

下图显示了连接到多启始端 SCSI 总线上的 JBOD 存储器内的终止情况。



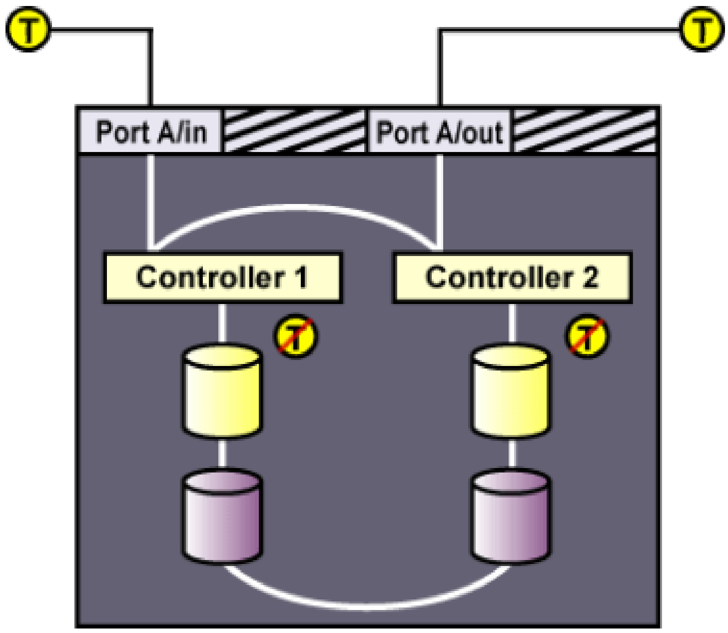
连接到多启始端 SCSI 总线的 JBOD 存储器

下图显示了连接到多启始端 SCSI 总线的单控制器 RAID 阵列内的终止情况。




连接到多启始端 SCSI 总线的单控制器 RAID 阵列

下图显示了连接到多启始端 SCSI 总线的双控制器 RAID 阵列内的终止情况。




连接到多启始端 SCSI 总线的双控制器 RAID 阵列

### 2.4.4.2 设置单启始端 SCSI 总线

单启始端 SCSI 总线仅能连接一个集群系统，可提供主机隔离能力与比多启始端总线更高的性能。它可确保每个集群系统免受由于另一个集群系统的工作负载、初始化或系统维修所带来的影响。

如果您使用的是带有多个主机端口的单控制器或双控制器 RAID 阵列，它能提供从这些端口到所有共享逻辑单元的并发访问，那么您可以设置两条单启始端总线或互连来将每个集群系统都连接到 RAID 阵列上。如果一个逻辑单元可以从一个控制器故障切换到另一个控制器，该过程对于操作系统来讲必须保持透明。

我们强烈推荐生产环境中使用单启始端SCSI总线或者单启始端光纤通道互连。

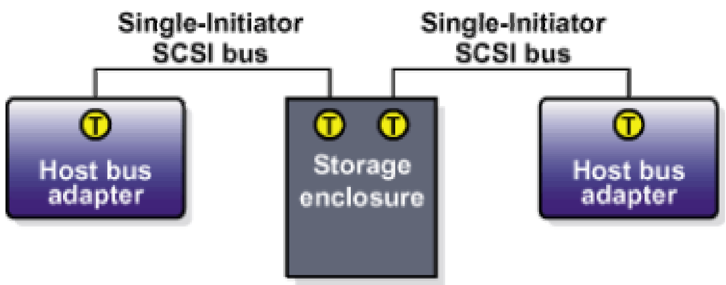
注意某些 RAID 控制器会将某组磁盘限定在特定的控制器或者端口上。这种情况下，您不能设置单启始端总线。此外，热插拔对于单启始端 SCSI 总线来讲不是必需的，因为专用总线无需在您从总线上断开主机总线适配器时依然保持可用。

单启始端总线必须满足 **SCSI 总线配置要求**中所描述的各项要求。另外，请参见主机总线适配器的特性与配置要求，了解有关终止主机总线适配器和配置单启始端总线的更多信息。

要设置单启始端 SCSI 总线，你必须遵照以下步骤进行：

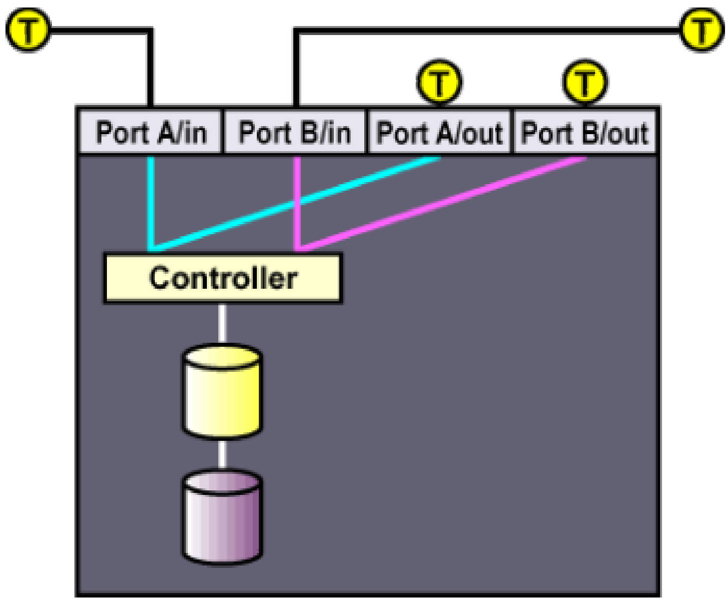
- 启用每个主机总线适配器的板载终端。
- 启用每个 RAID 控制器的终端。
- 使用适当的 68-针 SCSI 电缆将各主机总线适配器连接到存储器。

要设置主机总线适配器终端，您必须在系统启动过程中进入 BIOS 工具。要设置 RAID 控制器终端，请参见厂商文档。下图为使用两条单启始端 SCSI 总线的系统配置



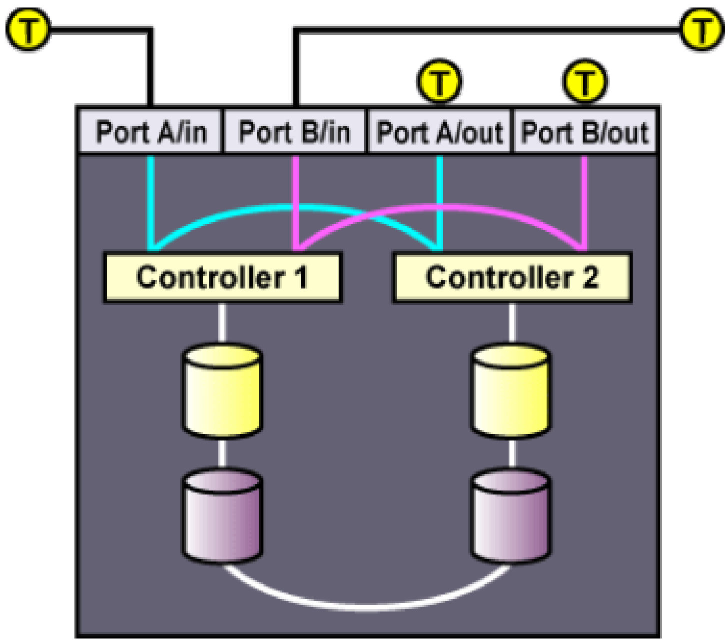

单启始端 SCSI 总线配置

下图为连接到两条单启始端 SCSI 总线的单控制器 RAID 阵列内的终止情况。




连接到单启始端 SCSI 总线的单控制器 RAID 阵列

下图为连接到两条单启始端 SCSI 总线的双控制器 RAID 阵列中的终止情况。




连接到单启始端 SCSI 总线的双控制器 RAID 阵列

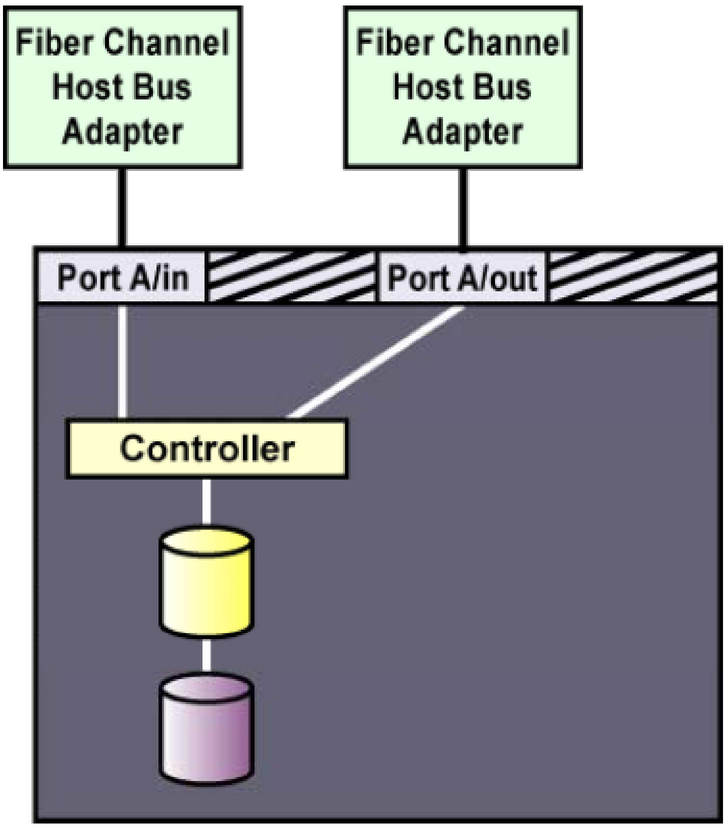
#### 2.4.4.3 设置单启始端光纤通道互连

单启始端光纤通道互连仅连接一个集群系统，可提供主机隔离能力与比多启始端总线更高的性能。它可确保每个集群系统免受由于另一个集群系统的工作负载、初始化或系统维修所带来的影响。

我们建议在产品环境中使用单启始端 SCSI 总线或者单启始端光纤通道互连。

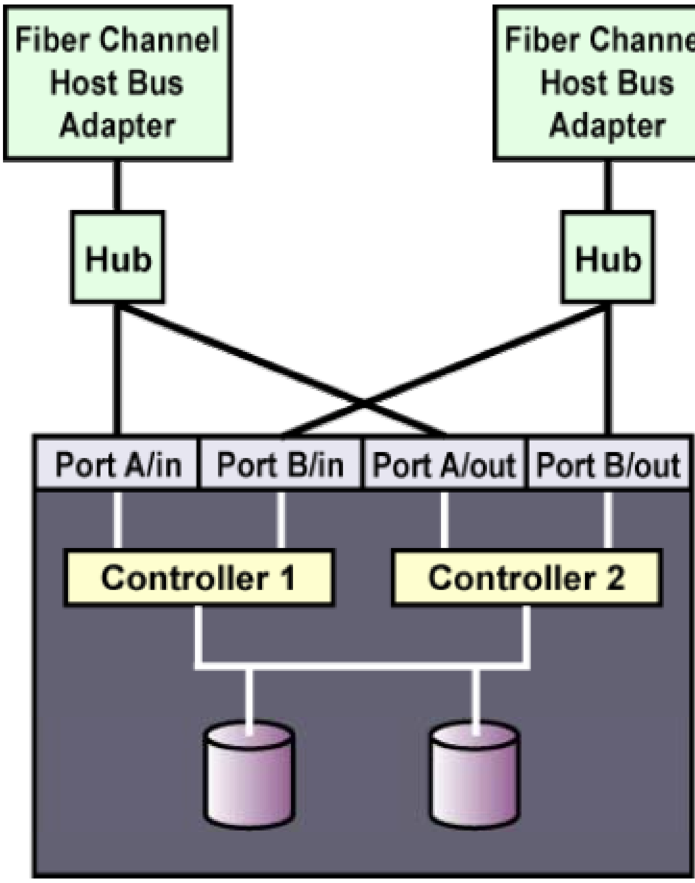
如果您的 RAID 阵列具有多个主机端口，并能提供从这些端口到所有共享逻辑单元的并发访问，那么您可以设置两条单启始端光纤通道互连来将每个集群系统都连接到 RAID 阵列上。如果一个逻辑单元可以从一个控制器故障切换到另一个控制器，该过程对于操作系统来讲必须保持透明。

下图中显示的单控制器 RAID 阵列带有两个主机端口，主机总线适配器直接连接到 RAID 控制器上，没有使用光纤通道总线或者交换机。




连接到单启始端的单控制器 RAID 阵列

如果您使用的是双控制器 RAID 阵列，每个控制器上都有 2 个主机端口，那么您必须使用一条光纤通道总线或者交换机将每个主机总线适配器连接到两个控制器的端口上，如下图所示。




连接到单启始端光纤通道互连的双控制器 RAID 阵列

#### 2.4.4.4 创建文件系统

使用 mkfs 命令来在分区上创建 ext2 文件系统，指定驱动器标识符和分区编号。例如：

```
# mkfs /dev/sde3
```

为了获得最佳的性能，在创建共享文件系统的时候请使用 4 KB 的块大小。注：某些 mkfs 文件系统创建程序默认的是 1 KB 块大小，这可能会导致较长的 fsck 时间。

要创建 ext3 或 reiserfs 文件系统，请使用 mkfs.ext3 和 mkfs.reiserfs 命令，而不是 mkfs。在 GreatTurbo HA 中，我们推荐使用日志文件系统，因为它们不需要 fsck，因而大大缩短了故障切换时间。Turbolinux Servers 对 ext3 和 reiserfs 文件系统都可支持。

### 2.4.5 配置磁盘镜像存储器

磁盘镜像存储与共享磁盘存储的功用相同，提供了一款高可用性的低成本解决方案。要构建磁盘镜像存储设备，您仅需在每个节点上准备一个磁盘分区，并在两个节点之间建立以太网通信即可。在此我们推荐使用网络交叉电缆将两个系统的网络接口连接起来，以提供不间断的高速数据传输。磁盘镜像存储的 I/O 性能取决于集群系统的网络接口。设备搭建好之后，所有操作都同共享磁盘存储方案一样。

# 第三章　安装 GreatTurbo Cluster Server 10

当完成集群硬件和操作系统的安装和配置之后，您必须安装 GreatTurbo Cluster Server 10 并进行系统的配置。

GreatTurbo Cluster Server 10 的安装将把 GreatTurbo Cluster、Agent、drbd 以及可能需要的附加支持包安装到系统中，安装需要大约 10MB 的磁盘空间。

## 3.1 安装 GreatTurbo Cluster Server 10

安装 GreatTurbo Cluster Server 10 的过程如下：

1) 确认您所使用的 GreatTurbo Cluster Server 10 产品的功能级别
2) 安装 GreatTurbo Cluster Server 10
3) 安装 guiadmin 客户端

### 3.1.1 确认您所使用的 GreatTurbo Cluster Server 10 产品的类型

GreatTurbo Cluster Server 10 产品分为两个功能级别，请确认您购买的产品属于哪一个级别。

- 第一级别只支持使用磁盘阵列的应用。
- 第二级别既支持使用磁盘阵列的应用，也支持 drbd（磁盘镜像设备）。

选购何种级别功能的 GreatTurbo Cluster Server 10，完全取决于您需要满足的应用类型。确定了 GreatTurbo Cluster Server 10 的功能级别之后，就可以开始安装 GreatTurbo Cluster Server 10 了。

### 3.1.2 安装 GreatTurbo Cluster Server 10

请分别在节点 A 和节点 B 安装 GreatTurbo Cluster Server 10 软件。

1） 如果您购买的是正版的 GreatTurbo Cluster Server 10 产品，GreatTurbo Cluster Server 10 会附带软件的安装光盘。插入安装光盘到节点的光驱并 mount 后，光盘根目录中有一个安装文件 install_cluster，请运行如下命令进行安装。根据您的操作系统版本并参照其提示完成 GreatTurbo Cluster Server 10 的安装。

如果您的 GreatTurbo Cluster Server 10 产品支持 drbd（磁盘镜像设备），请在“Do you want to use drbd ?”时选择 yes，否则选择 no。

例如：

```
[root@test1 root]# mount   /dev/cdrom   /mnt/cdrom
[root@test1 root]# cd   /mnt/cdrom
[root@test1 cdrom]# ./install_cluster
Following RPMs will be installed or upgraded to newer version if necessary:
   *) pdksh
   *) lsof
   *) sg_utils
   *) drbd
   *) GreatTurbo Cluster Server
   *) GreatTurbo Cluster Server Agents
Please select the operatating system version you are using:
   0) GreatTurbo Enterprise Server 10 (x86_32)
   1) GreatTurbo Enterprise Server 10 (x86_64)
   2) GreatTurbo Enterprise Server 10 (OpenPower ppc64)
   3) GreatTurbo Enterprise Server 10 (IA64)
   4) cancel
Please select an OS version, then the installation procedure will begin. Select 4 to abort
[0/1/2/3/4]: 0
Do you want to use drbd? (y/n) [y]: y
Please select the network forwarding method you want to use:
   0) direct routing
   1) NAT
```

```
  2) IP tunnel
Please select a network forwarding method [0/1/2]: 1
```

2） 如果您是从拓林思公司下载的试用产品（使用有效期为 1 个月），您下载的 GreatTurbo Cluster Server 10 将是一个 ISO 文件和 md5 检验文件。ISO 文件的名称一般为 greatturbocluster-10.x-x.iso，md5 文件的名称一般为 greatturbocluster-10.x-x.iso.md5。为了验证您下载的 iso 文件是否正确，可以先对您的 iso 文件进行 md5sum 操作，如果发现 md5sum 运行命令所得结果，和您下载的 md5 文件的内容一致，则表明您下载的 ISO 文件正确无误。

例如：

```
[root@test1 iso]# md5sum ./ greatturbocluster-10.0-1.iso
88f7b387c8cdaca3fa38832c6b7b2b6a  greatturbocluster-10.0-1.iso
[root@test1 iso]# cat ./ greatturbocluster-10.0-1.iso.md5
88f7b387c8cdaca3fa38832c6b7b2b6a  greatturbocluster-10.0-1.iso
```

证明您下载的 ISO 文件正确之后，请先将 ISO 文件 mount 到某一个目录，然后运行这个目录下的 install_cluster 安装程序进行安装，接下来的过程同从光盘安装时介绍的一样。

```
[root@test1 root]# mount –o loop   ./ greatturbocluster-10.0-1.iso   /mnt/iso
[root@test1 root]# cd   /mnt/iso
[root@test1 iso]# ./install_cluster
```

### 3.1.3 安装 guiadmin 客户端

如果用户仅仅在 GreatTurbo Cluster 的两个成员节点上运行 guiadmin 客户端软件，那么安装 GreatTurbo Cluster Server 10 即可，不需要做额外的操作。如果用户需要远程操作，那么需要在远程客户机上执行安装光盘里的 install_guiclient。

例如：

```
[root@test1 root]# mount   /dev/cdrom   /mnt/cdrom

[root@test1 root]# cd   /mnt/cdrom

[root@test1 cdrom]# ./install_guiclient
```

```
The GreatTurbo HA guiadmin client will be installed or upgraded to newer version.

Do you want to continue? (y/n) [y]: y

Preparing packages for installation...

GreatTurbo HA-guiadminclient-10.0-8

GreatTurbo HA guiadmin client installation finished.
```

## 3.2 注册 GreatTurbo Cluster Server 10

安装完 GreatTurbo Cluster Server 10 的软件之后，您需要注册 GreatTurbo Cluster Server 10 产品，也就是说需要安装 GreatTurbo Cluster Server 10 的 license，以保证 GreatTurbo Cluster Server 10 正常运行。

注册 GreatTurbo Cluster Server 10 产品的步骤如下：

1) 获得两个节点的硬件号，分别在两个节点上运行"/opt/cluster/bin/syncd -l"，其输出的第一行信息，例如"Hardware ID: 3355DEJGQWVK"，其中"3355DEJGQWVK"就是硬件号。仔细记录两台节点的硬件号，以便在随后的注册中使用。
2) 用浏览器登陆注册网站 http://www.turbolinux.com.cn/register。进入如下图所示的登陆页面：



3)　如果您是第一次注册，请点击“新用户”按钮，按要求详细填写注册信息。

4)　注册完之后，返回如上登陆页，输入电子邮件地址和密码，点击“登陆”，进入注册页。您还可以在产品序列号栏，输入您购买产品时附带的序列号，以查看序列号有没有被注册过，如果发现被人盗用注册过，则请与拓林思软件有限公司联系，因为注册码只允许注册一次。

5)　登陆后，进入以下页面：



如果您是正版用户，请点击“增加产品序列号”按钮、进入步骤 6）。如果您没有序列号，只是试用，请点击“评估版许可证”按钮、进入步骤 7）。

6)　点击“增加产品序列号”之后，进入如下页面：



7) 在 SN 栏输入您购买产品时附带在注册卡上的序列号、然后点击“下一步”按钮，或者点击 5）中“评估版许可证”进入如下页面：




8) 选择您的产品类型，点击“下一步”按钮，进入以下页面：




9) 分别填入两个节点的硬件号，然后点击完成，进入到如下所示的页面：




10) 点击“取得使用许可证”按钮，在接下来的页面上选择如何保存你得许可证：寄到信箱或者存到文件。
11) 得到 license 文件之后，你需要把它分别 copy 到两个节点的/etc/opt/cluster/lic 目录下。

**注意：**
**<u>在/etc/opt/cluster/lic 目录下只能保存一个 license 文件。</u>**

至此，GreatTurbo Cluster Server 10 产品注册完毕。接下来在第三章中，我们将讲述如何对 GreatTurbo Cluster Server 10 进行初始配置。

# 3.3 升级 GreatTurbo Cluster Server 10

## 3.3.1 GreatTurbo Cluster server 10 的升级

如果您希望升级 GreatTurbo Cluster Server 10 到最新版本，同时又希望保存现有的 GreatTurbo Cluster Server 的 配置文件。并且您想在更新 GreatTurbo Cluster Server 10 的同时最大限度地减少服务停机时间，则请遵循以下步骤进行操作：

(1) 在您想更新的集群系统的第一个节点上，运行 cluadmin 工具并备份当前的集群配置文件。例如：

```
cluadmin> cluster backup
```

关于 cluadmin 工具的使用方法，请参阅相关的章节。

(2) 将在第一个节点上运行的服务切换到第二个节点上。切换服务的方法，请参阅相关的章节。

(3) 通过调用 System V init 目录下的 cluster stop 命令来停止第一个节点上的 GreatTurbo Cluster Server 10。例如：

```
/etc/init.d/cluster stop
```

(4) 按照 GreatTurbo Cluster Server 10 的安装步骤中的说明，在您希望更新的第一个节点上安装最新的 GreatTurbo Cluster Server 10。

(5) 通过调用 System V init 目录下的 cluster stop 命令来停止第二个节点上的 GreatTurbo Cluster Server 10。此时，将没有任何服务可用。

(6) 通过调用 System V init 目录下的 cluster start 命令来启动第一个被更新的节点上的 GreatTurbo Cluster Server 10。此时，所有服务将变得可用。

(7) 按照 GreatTurbo Cluster Server 10 的安装步骤中的说明，在您希望更新的第二个节点上安装最新的 GreatTurbo Cluster Server 10。

(8) 通过调用 System V init 目录下的 cluster start 命令来启动第二个被更新节点上的 GreatTurbo Cluster Server 10。

至此，两个节点的 GreatTurbo Cluster Server 10 升级完毕。

# 第四章　卸载 GreatTurbo Cluster Server 10

卸载 GreatTurbo Cluster Server 10 的过程如下：

1) 卸载 GreatTurbo Cluster Server 10
2) 卸载 drbd
3) 卸载 GreatTurbo Cluster Server10 realserver 包
4) 卸载 guiadmin 客户端

## 4.1 卸载 GreatTurbo Cluster Server 10

您可以利用 GreatTurbo Cluster Server 10 提供的 uninstall_cluster 工具来卸载 GreatTurbo Cluster Server 10。uninstall_cluster 安装在/sbin 目录下。

**注意：**

**卸载的操作请在两个节点上分别执行。**

卸载的方法如下：

1) 卸载之前请先停止 GreatTurbo Cluster Server 10。GreatTurbo Cluster Server 10 停止的方法如下：

```
/etc/init.d/cluster stop
```

例如：

```
[root@test1 root]# /etc/init.d/cluster stop
-------- Performing cluster stop --------------------
---- Performing guiadmin server stop ----
Stopping guiadmin server: done.
---- Completed guiadmin server stop ----
Shutting down clumon:done
Sending stop message to svcmgr: done.
```

```
Waiting for Cluster Daemons to exit.
Stopping syncd: done.
Stopping hb: done.
Stopping svcmgr: done.
Stopping powerd: done.
Stopping svccheck: done.
____ Performing drbd stop ____
drbd: 'drbd0' already Secondary
drbd: module has been unloaded
____ Completed drbd stop ____
-------- Completed cluster stop ---------------------
```

2) 执行卸载工具 uninstall_cluster 来卸载 GreatTurbo Cluster Server 10。如果您安装了 drbd，请在提示是否卸载 drbd 时选择 y。

```
[root@test1 root]# ./uninstall_cluster
Following RPMs will be uninstalled:
   *) GreatTurbo Cluster Server
   *) GreatTurbo Cluster Server Agents
   *) drbd(if installed)
Do you want to continue? (y/n) [y]: y
Do you want to uninstall drbd? (y/n) [y]: y
GreatTurbo Cluster Server uninstallation finished.
```

## 4.2 卸载 drbd

如果您安装了 drbd，执行 uninstall_cluster 会卸载 drbd 软件包。如果您要单独卸载 drbd，请进行以下操作：

(1) Linux kernel2.4 环境

在 Linux kernel2.4 环境下，卸载 drbd 软件包，请执行以下的命令：

```
[root@test1 root]# rpm -e drbd
```

(2) Linux kernel2.6 环境

在 Linux kernel2.6 环境下，卸载 drbd 软件包，请执行以下的命令：

其中，xxx 代表 Linux kernel 版本号。

```
[root@test1 root]# rpm -e drbd-km-xxx

[root@test1 root]# rpm -e drbd
```

## 4.3 卸载 GreatTurbo Cluster Server10 realserver 包

卸载的方法如下：

1) 卸载之前请先恢复 realserver 配置，根据选择的负载技术执行不同的命令：

```
[root@test1 root]# /etc/init.d/lbrealserver-dr stop
```

2) 执行卸载工具 uninstall_realserver 来卸载 GreatTurbo Load Balance Server-realserver

```
[root@test1 root]# uninstall_realserver
```

## 4.4 卸载 guiadmin 客户端

执行 install_cluster 在 GreatTurbo Cluster Server 10 集群的两个节点上安装的 guiadmin 客户端，在执行 uninstall_cluster 时会自动卸载。

如果在别的机器上执行 install_guiclient，单独安装了 guiadmin 客户端，请执行以下的命令进行卸载：

```
rpm -e greatturbocluster-guiadminclient
```

# 第五章　配置 GreatTurbo Cluster Server 10

## 5.1 member_config 工具说明

member_config 是用来对 GreatTurbo Cluster Server 10 进行初始化配置的工具。member_config 工具主要完成以下的工作：

- 配置集群的节点名称
- 配置集群的心跳
- 配置电子开关
- 配置 watchdog
- 配置第三方参考 IP
- 创建配置文件

您需要如下信息才能使用 member_config 来初始化集群。这些信息要输入到 GreatTurbo Cluster Server10 配置文件的成员字段中，您可以在 /etc/opt/cluster/cluster.conf 文件中找到这些成员字段：

- 通过 hostname 命令返回的集群系统主机名
- 心跳连接（通道）的数量，包括以太网和串行连接
- 每个心跳串行连接的设备文件，如：/dev/ttyS1
- 同每个心跳以太网接口相对应的 IP 主机名称
- 与电子开关连接的串行口的设备文件，如：/dev/ttyS0

## 5.2 配置 GreatTurbo Cluster Server 10

对 GreatTurbo Cluster Server 10 进行初始化配置的过程如下：

(1) 选择其中一节点（例如节点 A）进行系统配置；

(2) 在另一节点（例如节点 B）同步配置信息。

**注意：**

**member_config 只需要在一个节点上配置。**

### 5.2.1 选择其中一节点进行初始化配置

首先我们需要选择其中一节点对 GreatTurbo Cluster Server 10 进行一些配置。

1）启动 member_config 命令

**注意：**

如果您使用了 IBM 的 EXP400 系列的磁盘柜，在确定 IBM 的 EXP400 磁盘阵列的硬件安装和配置正确之后，运行 member_config -s 命令。

运行 member_config 命令，系统显示如下：

```
[root@test1 root]# /opt/cluster/bin/member_config

------------------------------------

Cluster Member Configuration Utility

------------------------------------

Version: 10.0 Built: Fri Jun 9 16:51:41 CST 2006


This utility sets up the member systems of a 2-node GreatTurbo Cluster Server 10 cluster,

or the 2-node director members of a Load Balancing cluster.

It prompts you for the following information:
```

```
o    Hostname
o    Number of heartbeat channels
o    Information about the type of channels and their names
o    Power switch type and device name
o    Information about the routers and network type of the Load Balancer

In addition, it performs checks to make sure that the information
entered is consistent with the hardware, the Ethernet ports, the raw
partitions and the character device files.

After the information is entered, it initializes the configure file
and saves the configuration information to the configure file

- Checking that cluster daemons are stopped: done
- Configuration file exists already.
   Would you like to use those prior settings as defaults? (yes/no) [yes]:
```

如果以前运行过 member_config 命令，那么配置文件中会保存我们的配置结果，该选项就是问我们是否使用以前的配置结果作为缺省值，通常回答 yes。

**注意：<u>如果是第一次配置，则不会出现该选项。</u>**

## 2）配置本地节点名称

接下来，输入本地节点的名称。GreatTurbo Cluster Server 10 会自动的从您系统中/etc/hosts 中得到本地节点的名称，前提是您没有把这个名称对应到 127.0.0.1。 如果这个名称对应到 127.0.0.1，安装程序会退出，并提醒您正确的配置/etc/hosts。

```
----------------------------------------
```

```
Setting information for cluster member 0
----------------------------------------
Enter name of cluster member [test1]: test1
Looking for host test1 (may take a few seconds)...
Host test1 found
Cluster member name set to: test1
```

### 3）配置本地节点的 heartbeat 选项：

GreatTurbo Cluster Server 10 的 heartbeat 通道有两类： 网络(net)和串口(serial)。对于网络，您需要配置 heartbeat 使用的网络设备对应的别名；对于串口，您需要配置 heartbeat 使用的串口的设备名，例如 /dev/ttyS0。

**注意：**<u>为了获得更高的可用性，GreatTurbo Cluster Server 10 建议配置一条串口通道以及至少两条直连网络心跳通道，并且必须将应用所在的网卡配置成通道，配置通道的顺序为：先配置所有的直连网线通道，再配置应用所在网卡的通道，最后配置串口通道。</u>

```
Enter number of heartbeat channels (minimum = 1) [1]: 4
You selected 4 channels
Information about channel 0:
Channel type: net or serial [net]:
Channel type set to: net
Enter hostname of cluster member test1 on heartbeat channel 0 [test1]: hb11
Looking for host hb11 (may take a few seconds)...
Host hb11 found
Hostname corresponds to an interface on member 0
Channel name set to: hb11
Information about channel 1:
Channel type: net or serial [net]:
```

```
Channel type set to: net
Enter hostname of cluster member test1 on heartbeat channel 1: hb12
Looking for host hb12 (may take a few seconds)...
Host hb12 found
Hostname corresponds to an interface on member 0
Channel name set to: hb12
Information about channel 2:
Channel type: net or serial [net]:
Channel type set to: net
Enter hostname of cluster member test1 on heartbeat channel 2: test1
Looking for host test1 (may take a few seconds)...
Host test1 found
Hostname corresponds to an interface on member 0
Channel name set to: test1
Information about channel 3:
Channel type: net or serial [net]: serial
Channel type set to: serial
Enter device name: /dev/ttyS0
Device /dev/ttyS0 found and no getty running on it
Device name set to: /dev/ttyS0
```

### 4）配置本地节点的 power switch 和 watchdog 选项：

如果没有硬件电子开关设备，请输入 NONE。如果有硬件电子开关，输入相应的电子开关对应的类型，如：RSA、RPS10 或者是 APC。

软件级的 watchdog 可以用来保证 GreatTurbo Cluster Server 10 程序的健壮性但并不能保障操作系统的自动重启恢复； 如果有硬件 watchdog，请输入硬件

watchdog 对应的驱动模块的名字,如果没有硬件 watchdog,建议配置软件 watchdog,输入操作系统默认附带的软件 watchdog 的模块名字 softdog 即可。

```
Information about power switch connected to member 0
Specify one of the following switches (NONE/RSA/RPS10/APC) [NONE]: NONE
Power switch type set to NONE

Information about watchdog to member 0
Choose one of the following watchdog drivers: NONE/softdog/...) [NONE] : softdog
```

5）配置对方节点信息

对方节点信息包括节点的机器名，heartbeat 的设置，watchdog driver 等。设置方法和本地完全一样。

```
----------------------------------------
Setting information for cluster member 1
----------------------------------------
Enter name of cluster member: test2
Looking for host test2 (may take a few seconds)...
Host test2 found
Cluster member name set to: test2

You previously selected 4 channels
Information about channel 0:
Channel type selected as net
Enter hostname of cluster member test2 on heartbeat channel 0: hb21
Looking for host hb21 (may take a few seconds)...
```

```
Host hb21 found
Channel name set to: hb21
Information about channel 1:
Channel type selected as net
Enter hostname of cluster member test2 on heartbeat channel 1: hb22
Looking for host hb22 (may take a few seconds)...
Host hb22 found
Channel name set to: hb22
Information about channel 2:
Channel type selected as net
Enter hostname of cluster member test2 on heartbeat channel 2: test2
Looking for host test2 (may take a few seconds)...
Host test2 found
Channel name set to: test2
Information about channel 3:
Channel type selected as serial
Enter device name [/dev/ttyS0]: /dev/ttyS0
Device name set to: /dev/ttyS0

Information about power switch connected to member 1
Specify one of the following switches (NONE/RSA/RPS10/APC) [NONE]: NONE
Power switch type set to NONE


Information about watchdog to member 1
Choose one of the following watchdog drivers(NONE/softdog/...) [NONE]: softdog
```

## 6）配置磁盘心跳设备

利用共享 raw 磁盘分区作为心跳通道后，只要主备节点能够访问共享数据，就不会发生裂脑，从而有效的确保了共享数据的一致性。如果需要配置磁盘心跳，请选择 yes 继续配置。如果条件实在不具备，输入 no，然后回车。

```
---------------------------
Setting up raw disk heartbeat
---------------------------

Do you want add raw disk heartbeat device? (yes/no) [yes]:
The raw disk heartbeat device must be a raw disk device.
To enhance redundancy, need to configure two raw disk devices.

Enter the name of first raw disk heartbeat device [/dev/raw/raw1]: /dev/raw/raw1

Enter the name of second raw disk heartbeat device [/dev/raw/raw2]:
/dev/raw/raw1
Now begin to initialize the raw disk heartbeat device, it will cause the data in the devices lost.
Are you sure to initialize the raw devices which you input? (yes/no) [yes]:yes
```

## 7）配置第三方 IP 地址

您需要配置两个节点都可以连接的第三方 IP 地址（要求能 ping 通，一般选择网关作为第三方 IP）。第三方 IP 能配置多个，以进一步提高系统高可用性。**如果条件具备，建议您配置第三方 IP。**

如果条件实在不具备，输入 no，然后回车。

```
------------------------------
Setting up the third partner ip
```

```
------------------------------

Do you want cluster to determine network status? (yes/no) [yes]:

The IP address of third computer is needed to determine network status.
The third computer should be up all the time, so it is recommend to use
gateway IP address here. Please use IP address instead of domain name.
If you want input multiple third part IP, please use comma to separate.
For example, 192.168.0.1,192.168.0.3,192.168.0,10

Enter the IP address of third computer [172.16.76.1]:
    ……
```

## 8）配置负载均衡调度信息

首先需要输入两个调度节点的 ip 地址，然后选择一种调度方式。注意此时选择的调度方式必须和安装软件包时所选择的方式一致（请选择和 realserver 保持一致的调度方式）。如果选择 NAT 作为负载调度技术，那么还需要配置 NAT router 地址的相关信息。

```
---------------------------------------
Setting Load Balancer informations
---------------------------------------
Enter load balancer primary server IP [172.16.70.138]:
Looking for IP address 172.16.70.138 (may take a few seconds)...
IP address 172.16.70.138 found
IP address corresponds to an interface on member 0
Load balancer primary server IP set to: 172.16.70.138
```

```
Enter load balancer backup server IP [172.16.70.76]:
Looking for IP address 172.16.70.76 (may take a few seconds)...
IP address 172.16.70.76 found
Load balancer backup server IP set to: 172.16.70.76
Enter load balancer network forwarding type(direct/nat/tunnel) [direct]: nat
Load balancer network forwarding type set to: nat
Enter NAT router IP address: 172.16.70.76
NAT router IP address set to: 172.16.70.76
Enter NAT router netmask: 255.255.255.0
NAT router netmask set to: 255.255.255.0
Enter NAT router device(e.g. eth1:1): eth1:1
NAT router device set to: eth1:1
------------------------------------
The following choices will be saved:
------------------------------------
        ……
```

### 9）保存配置

如果以上内容全部配置完毕，member_config 将会询问是否保存改动。如果刚才的配置没有错误，请输入"yes"或直接回车；

然后程序还会询问是否运行"diskutil -I"来初始化配置文件，这里请选择"yes"。

**注意：当 GreatTurbo Cluster Server 10 的守护进程正在运行时，请不要选择运行 diskutil -I 来初始化配置文件，以免造成不可预知的后果。**

```
Save changes? yes/no [yes]:
Writing to output configuration file...done.
Changes have been saved to /etc/opt/cluster/cluster.conf.
```

```
---------------------------
Setting up Configure File
---------------------------
Run diskutil -I to set up the configure file now?
- Select 'yes' to clean up a previous install
- Select 'no' if you have just set them up on other member
    and have not started the cluster services on that member
- Select 'no' if you are running it on other cluster member
Warning: Do not run 'diskutil -I' on a running cluster, because it would have severe consequences.
yes/no [no]: yes

Saving configuration information to configure file: done


--------------------------------------------------------------
Setup on this member is complete.    If errors have been reported,
correct them.

If you have not already set up the other cluster member, before
running member_config, invoke the following operation on the
other cluster member:
copy /etc/opt/cluster/cluster_raw.conf to another node with the same path.
cluster daemons on each cluster member by invoking the cluster start
script located in the System V init directory.    For example:
# /etc/rc.d/init.d/cluster start
```

### 5.2.2 在对方节点上同步配置

配置完成之后，需要把配置文件/etc/opt/cluster/cluster_raw.conf 手工拷贝到另外一台机器的相同目录里(使用 scp 或者 ftp 均可)。

```
[root@test2 root]# scp test1:/etc/opt/cluster/cluster* /etc/opt/cluster
```

### 5.2.3 利用备份的配置文件配置 GreatTurbo Cluster Server 10

如果您想使用过去备份的配置文件 cluster.conf 来对 GreatTurbo Cluster Server 10 进行初始化配置，您可以使用下面简单的命令在其中一台节点上生成配置文件。

```
[root@test2 root]# /opt/cluster/bin/diskutil –l
[root@test2 root]# /opt/cluster/bin/clu_config –f cluster.conf
```

然后您可以将配置文件/etc/opt/cluster/cluster_raw.conf 手工拷贝到另外一台节点的相同目录里。

**注意：**

**<u>请不要手工编辑cluster.conf 文件，请使用cluadmin 工具或者guiadmin工具来修改文件。</u>**

GreatTurbo Cluster Server 10 集群的初始化配置完成以后，您就可以添加服务了。请参见服务配置与管理的相关章节来了解更多信息。

## 5.3 运行 GreatTurbo Cluster Server 10

初始化完成之后，需要在两边节点分别运行 GreatTurbo Cluster Server 10。脚本/etc/init.d/cluster 可以用来启动 GreatTurbo Cluster Server 10。

```
[root@test1 root]# /etc/init.d/cluster start
```

```
-------- Performing cluster start ---------------------
Starting Turbo cluster...done.
---- Performing guiadmin server start ----
Starting guiadmin server...done.
---- Completed guiadmin server start ----
-------- Completed cluster start ---------------------
```

注意：<u>请不要执行完/etc/init.d/cluster start 后立即执行/etc/init.d/cluster stop，请确认 GreatTurbo Cluster Server 10 启动完成后再执行/etc/init.d/cluster stop。</u>

member_config 初始化完成之后，GreatTurbo Cluster Server 10 就可以正常运行了，只不过没有配置服务。用户可以参阅第九章的检测方法来判断 GreatTurbo Cluster Server 是否可以成功启动。

## 5.4 停止 GreatTurbo Cluster Server 10

如果想停止 GreatTurbo Cluster Server 10,需要在两边节点分别运行脚本/etc/init.d/cluster stop。

```
[root@test1 root]# /etc/init.d/cluster stop
-------- Performing cluster stop ---------------------
---- Performing guiadmin server stop ----
Stopping guiadmin server: done.
---- Completed guiadmin server stop ----
Shutting down clumon: done.
Sending stop message to svcmgr: done.
Waiting for Cluster Daemons to exit.
Stopping syncd: done.
```

```
Stopping hb: done.
Stopping svcmgr: done.
Stopping powerd: done.
Stopping svccheck: done.
-------- Completed cluster stop ---------------------
[root@test1 root]#
```

注意：

**请不要执行完/etc/init.d/cluster start 后立即执行/etc/init.d/cluster stop，请确认 GreatTurbo Cluster Server 10 启动完成后再执行/etc/init.d/cluster stop。**

# 第六章 配置和管理工具说明

GreatTurbo Cluster Server 10 有两个配置和管理工具：文本界面的 cluadmin 工具和图形界面的 guiadmin 工具。

## 6.1 cluadmin 工具

cluadmin 是 GreatTurbo Cluster Server 10 的文本界面配置和管理工具，位于 /opt/cluster/bin 路径下。cluadmin 类似 bash，可以使用 TAB 键进行命令补全。

利用 cluadmin 工具，您可以完成以下的任务：

- 配置和管理集群
- 显示集群的状态
- 配置和管理服务
- 显示服务的状态
- 配置和管理磁盘镜像设备

cluadmin 工具使用 advisory lock（咨询锁）来防止 GreatTurbo Cluster Server 10 配置文件被多个用户同时修改。只有占有 advisory lock 的用户才可以修改配置文件。在您调用 cluadmin 工具的时候，GreatTurbo Cluster Server 10 将检查锁是否已经分配给别的用户了；如果没有分配出去，那么 GreatTurbo Cluster Server 10 就会把它分配给您。在您退出 cluadmin 工具时会自动释放锁。

如果别的用户正占有该锁，程序会显示警告提醒您配置文件数据库已被锁定。GreatTurbo Cluster Server 10 会让您选择是否强占该锁。如果您选择了强占该锁，之前的占有者将不能继续修改 GreatTurbo Cluster Server 10 的 配置文件。

如果不是迫不得已最好不要强占锁，因为不一致的同时配置可能会导致不可预料的集群行为。此外，我们建议每次只对 GreatTurbo Cluster Server10 配置文件进行一项改动（添加、修改或删除服务）。

您可以在 cluadmin 命令行中指定如下选项：

| 选项 | 说明 |
| --- | --- |
| -d or --debug | 显示详细的诊断信息，用于调试 |
| -n or --nointeractive | 绕过 cluadmin 工具的顶级命令循环处理，用于 cluadmin 的调试 |
| -s or --stack-trace | 显示 cluadmin 的 stack trace，用于调试 |
| -V or –version | 显示 cluadmin 的当前版本信息 |
| -h, -?, or – help | 显示工具的帮助信息 |

一般情况下，使用 cluadmin 不需要指定命令选项，直接执行 cluadmin 命令就可以了。例如：

```
[root@test1 root]# cluadmin
Fri Jul   8 10:47:49 CST 2005

You can obtain help by entering help and one of the following commands:

cluster        service             clear
help            apropos             nbd
exit
cluadmin>
```

**注意：**

**如果您使用的是 Turbolinux 中文版，则在运行 cluadmin 之前，请先在 bash 下执行“ unset LC_CTYPE ”。**

运行/opt/cluster/bin/cluadmin，然后在 cluadmin 中按下两次 TAB 键，会显示如下所示的所有命令。（[Tab] [Tab]表示连续按两下 Tab 键，有的系统可能只需要按一次 Tab 键）

```
cluadmin>[Tab] [Tab]
apropos
clear
exit
help
cluster status
cluster monitor
cluster loglevel
cluster loglevel syncd
cluster loglevel svcmgr
cluster loglevel svccheck
cluster loglevel powerd
cluster loglevel heartbeat
cluster loglevel clumon
cluster heartbeat
cluster mail from
cluster mail to
cluster mail smtpserver
cluster mail level
cluster watchdog
cluster reload
```

```
cluster name
cluster edit
cluster backup
cluster restore
cluster saveas
cluster restorefrom
service add
service show state
service show config
service show services
service modify
service disable
service enable
service relocate
service delete
nbd add
nbd delete
nbd show
help apropos
help clear
help exit
help help
help cluster status
help cluster monitor
help cluster loglevel
help cluster reload
help cluster name
help cluster edit
help cluster backup
```

```
help cluster restore
help cluster saveas
help cluster restorefrom
help service add
help service show state
help service show config
help service show services
help service modify
help service disable
help service enable
help service relocate
help service delete
help nbd add
help nbd delete
help nbd show
```

下表介绍了 cluadmin 工具的命令和子命令：

| 命令 | 子命令 | 说明 |
|---|---|---|
| apropos | — | 显示与指定字符串参数匹配的cluadmin 命令，如果没有指定参数则显示所有 cluadmin 命令。例如：<br>cluadmin> apropos service |
| clear | — | 清除屏幕显示。 |
| exit | — | 退出cluadmin。<br>其它的退出命令还可以有quit, q。 |
| help | — | 显示指定cluadmin 命令或子命令的帮助信息。例如：<br>cluadmin> help service add |
| cluster | status | 显示当前集群状态的快照。例如： |

| | | |
|---|---|---|
| | | cluadmin> cluster status |
| Cluster | monitor | 5 秒为间隔不断地显示集群状态快照。按下Return 或Enter 回车键即可停止显示。您可以加上带有数值参数的-interval 选项，程序将按照参数指定的时间间隔（以秒为单位）显示状态快照。此外，您可以加上带有yes 或no 参数的-clear 选项：Yes 表示每显示一次就清屏；no 表示不清屏。例如：<br>cluadmin> cluster monitor -clear yes -interval 10 |
| Cluster | loglevel | 显示当前各个进程的日志级别，加上参数可以为进程指定日志级别。例如：<br>cluadmin> cluster loglevel syncd 7 |
| Cluster | heartbeat | 设置心跳端口，间隔和tko_count。例如：<br>cluadmin> cluster heartbeat interval 20 |
| Cluster | mail | 设置邮件告警的源地址，目的地址，邮件服务器和发送邮件的事件的日志级别。例如：<br>cluadmin> cluster mail from name@turbolinux.com.cn |
| Cluster | watchdog | 设置 watchdog 的 timeout。例如：<br>cluadmin> cluster watchdog wdtimeout 600 |
| Cluster | reload | 强制各个进程重新读配置文件 |
| Cluster | name | 将集群的名称改为指定的名称。集群名称包括在clustat，cluster 监视命令的输出结果中。例如：<br>cluadmin> cluster name dbcluster |
| Cluster | backup | 备份配置文件，备份后的配置文件的名字为：/etc/opt/cluster/cluster.conf.bak |
| cluster | restore | 从备份的配置文件/etc/opt/cluster/cluster.conf.bak 恢复配置。在执行 cluster restore 之前，另一端的 |

| | | |
|---|---|---|
| | | GreatTurbo HA 必须停止。 |
| cluster | saveas | 以指定文件名的方式备份配置文件，例如：<br><br>cluadmin> cluster saveas<br>/etc/opt/cluster/cluster_backup.conf |
| cluster | restorefrom | 从指定的文件恢复 GreatTurbo HA 的配置。例如：<br><br>cluadmin> cluster restorefrom<br>/etc/opt/cluster/cluster_backup.conf |
| service | add | 添加服务。 |
| service | show state | 显示服务的状态。 |
| service | show config | 显示服务的配置。 |
| service | show services | 显示所有的服务。 |
| service | modify | 修改服务。 |
| service | disable | 禁用服务。 |
| service | enable | 启用服务。 |
| service | relocate | 切换服务。 |
| service | delete | 删除服务。 |
| Nbd | add | 添加镜像设备。 |
| Nbd | delete | 删除镜像设备。 |
| Nbd | show | 显示镜像设备的配置。 |

在运行 cluadmin 工具时，您可以使用 Tab 键来帮助确认 **cluadmin** 命令。

- 在cluadmin> 下按住Tab 键可显示所有命令列表。
- 在提示符后输入字符，按下Tab 键可显示以输入字符开头的命令。
- 输入一个命令，按下Tab 键可显示该命令包含的所有子命令。
- 此外，您可以使用上下箭头键在提示符处显示使用过的cluadmin 命令。

**注意：**

**因为 GreatTurbo Cluster Server 10 是通过网络同步配置，所以只有当两个节点的 GreatTurbo Cluster Server 10 同时运行，才可以使用配置工具来进行配置；或者在两个节点 GreatTurbo Cluster Server 10 启动之前，只在一个节点进行一些配置，然后在 H GreatTurbo Cluster Server 10 启动时由 GreatTurbo Cluster Server 10 自动同步。但是不可以在两个节点的 GreatTurbo Cluster Server 10 都没有启动时，在两个节点都运行配置工具，这样 GreatTurbo Cluster Server 10 因为无法判断哪个配置更新，而无法同步配置。另外 GreatTurbo Cluster Server 10 采用时间戳来判断两个节点的配置是否同步，所以不要轻易修改系统时间。如果确实需要修改系统时间，则需要手工 copy 其中一个节点的配置文件 /etc/opt/cluster/cluster_raw.conf 到另外一个节点上（这时候 GreatTurbo Cluster Server 10 应该处于停止状态），再重新启动 GreatTurbo Cluster Server 10。这种方法也适合所有配置冲突的情况。**

# 6.2 guiadmin 工具

## 6.2.1 guiadmin 介绍

GreatTurbo Cluster Server 10 提供了一个图形用户界面（guiadmin），可用于配置、管理和监控系统（guiadmin 目前不支持对 LB 服务的管理）。guiadmin 支持跨平台操作，并可以实现对系统的远程监控。guiadmin 分为客户端和服务器两部分，服务器端运行在 GreatTurbo Cluster Server 10 系统的两个节点上；客户端可以运行在任意操作系统上，如：linux、unix、windows 系列操作系统。由于 guiadmin

的客户端是一种 GUI 程序，所以如果您在 linux 或者 unix 系统使用，必须在客户端系统中先启动 X Windows，在 X Windows 下运行该客户端程序。

guiadmin 可以实现很多管理和配置的功能，其中包括：

- 配置、修改心跳参数
- 配置、修改集群进程日志级别
- 配置、修改邮件提示参数
- 配置、修改 watchdog 信息
- 显示节点的配置信息
- 增加服务
- 修改服务
- 禁用服务
- 启用服务
- 切换服务
- 删除服务
- 显示集群和服务状态

### 6.2.2 guiadmin 模块介绍

guiadmin 能够实现的功能与 cluadmin 非常相似，绝大部分在 cluadmin 中实现的功能在 guiadmin 里面均可以实现，并且使用起来更加方便，快捷。运行 guiadmin 程序时，您将看到带有两个选项卡的主窗口。每个选项卡代表一个模块，用户可以点击选项卡来进行模块切换：

1. GreatTurbo HA（配置）：该模块用来配置和监视各 GreatTurbo Cluster server 10 各组件和服务的相关信息。

2. Status（状态）：该面板显示了两个集群节点的当前状态。每 3 秒刷新一次。

配置模块分为左右两部分。左边是节点树，不同的节点代表不同的功能项；右边是主面版，显示每个功能项的具体信息。

节点树的底部有两个控制按钮。

1. “Apply”：把用户所做的全部修改提交到服务器端，不退出客户端界面。

2. “Close”：取消全部修改并退出。 如果用户对服务进行了“Remove”、“Enable”、“Disable”、“Relocate”操作，则点击“Cancel”键不能够恢复以上动作。

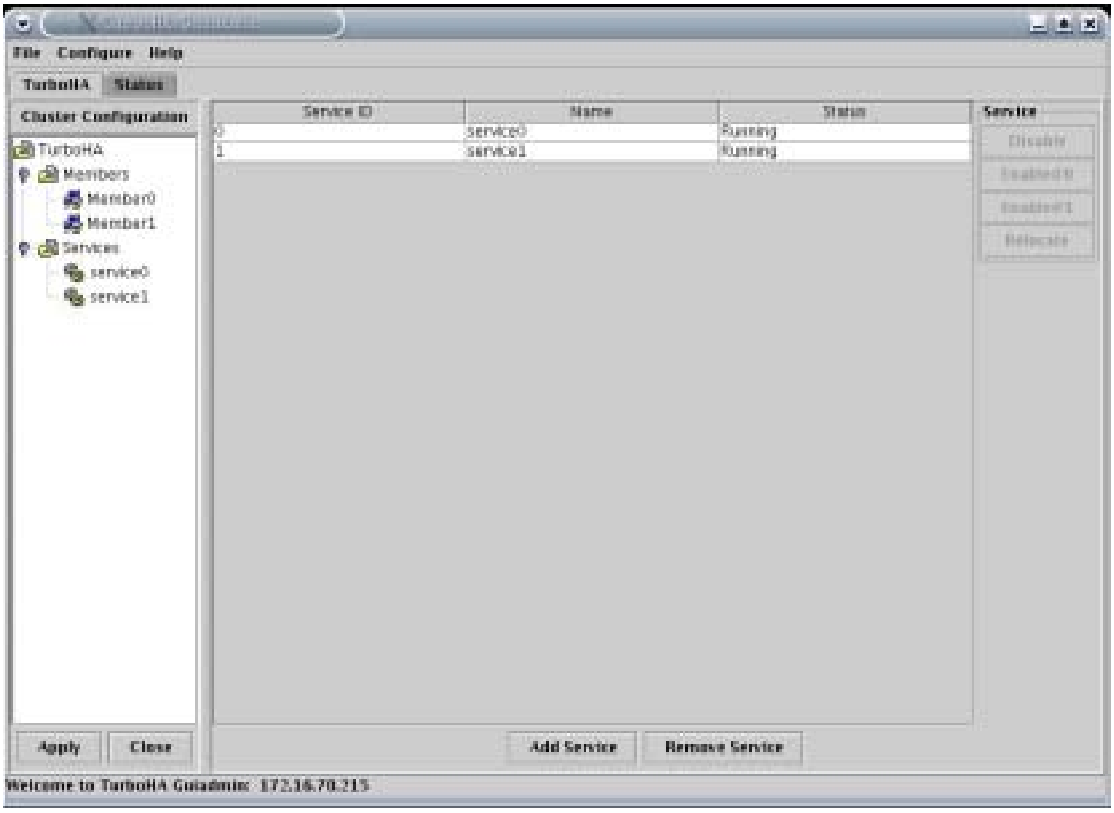
**注意：**

**1：guiadmin 工具每次只能在一个集群系统上运行，即不能够使用两个客户端程序同时对一个集群系统进行操作。**

**2：使用 guiadmin 工具的同时不要使用其它工具来进行配置。**

**3：使用 guiadmin 过程中如果 GreatTurbo Cluster Server 10 数据发生变化，或者服务状态发生变化，请重新启动客户端进程或者重新连接服务器。**

下图是已经和服务器建立连接的客户端主窗口：




下表描述了整个 guiadmin 中各个模块及窗口的特性。

| 特性 | 说明 |
|---|---|
| 配置模块 | 能够查看和修改所有集群和服务配置数据的核心管理工具。<br>可以查看和改变所有服务状态的核心服务管理工具。 |
| 状态模块 | 可以实时查看集群状态的监视工具。 |
| Apply（应用）按钮 | 提交配置模块中的条目或数据改动到服务器端，并使用户的改动生效。不关闭 guiadmin 客户端进程。 |
| Cancel（取消）按钮 | 取消配置模块中的数据条目，关闭 guiadmin 客户端进程。 |

### 6.2.3 登陆密码的设定

为了保证安全性，guiadmin 需要客户端使用者在连接服务器端的同时输入用户名和密码。用户名和密码由管理人员使用密码生成工具设置，/opt/cluster/bin/passwdgenerate 就是专门生成 guiadmin 连接密码的工具。

运行 passwdgenerate 命令，按照提示输入用户名和密码（用户名、密码不能够为空），这样用户名和密码就保存在磁盘上。当用户连接服务器端的时候，服务器端进程会检查用户名和密码的正确性。

如果用户想修改以前所设定的密码，那么可以再次运行该命令，按照提示输入用户名和密码即可。

用户在一个节点上设置完密码后,需要把/opt/cluster/lib/passwd 文件手动拷贝到另外一个节点的/opt/cluster/lib 目录下，保持两个节点的连接密码一致。

# 第七章 配置和管理 GreatTurbo Cluster Server 10 的服务

## 7.1 配置服务前的准备工作

如欲配置某项服务，您必须为其准备好所需的相应的资源信息。配置服务前的准备工作主要有：

1) 收集服务相关的信息。

2) 创建服务的启动和停止脚本。

3) 选择或编写供检测服务使用的应用代理。

4) 设置服务使用的磁盘存储设备。

### 7.1.1 收集服务信息

在创建某项服务之前，您必须收集有关的服务资源和属性信息。当您用 cluadmin 工具将某项服务添加到 GreatTurbo Cluster server 配置文件时，cluadmin 工具将向您做出信息提示。

在某些情况下，您可以为一项服务指定多个资源。例如，您可以指定多个 IP 地址和磁盘设备。

下表中对您可以指定的服务属性和资源进行了说明。

| 服务属性或资源 | 说明 |
| --- | --- |

| 服务属性或资源 | 说明 |
|---|---|
| 服务名称 | 每项服务都必须拥有一个单独的服务名称。每个服务名称可以包含 1－63 个字符，并且必须由字母（大写字母或小写字母）、整数、下划线、点和破折号等组成。但是，服务名称必须以字母或下划线开头。 |
| 首选成员 | 指定在不发生故障切换或者不通过手动对服务进行切换时，您希望在哪个集群节点（如果有的话）上运行该服务。 |
| 用户脚本 | 用来启动和停止服务的脚本。 |
| 检测脚本，检查的时间间隔，检查的timeout，允许的最大出错次数 | 检测脚本用来在服务运行过程中对服务的运行情况进行检查，服务检测进程每隔一定的时间间隔调用一次检测脚本，如果在指定的timeout 时间内检测脚本没有返回，则认为本次检查失败，如果服务检测出错的次数超过了允许的最大出错次数，则认为服务出错。 |
| 浮动 IP | 您可以为某项服务指定一个或多个浮动IP地址。对于某个集群系统来说，该IP地址与带有主机名称的以太网接口的IP地址不同，因为在出现故障时，它可随同服务资源一起自动重新定位。如果客户机使用该IP地址来访问服务，则它们不知道哪一台节点在运行服务，并且故障切换对于客户机是透明的。注意，集群节点必须在某项服务中指定的每一个IP地址的所在子网中配有网卡。您还应当为每一IP地址指定子网掩码和广播地址。 |
| 磁盘镜像设备 | 如果您的服务需要使用磁盘镜像设备，您必须指定磁盘镜像设备的 ID 号，这个 ID 号是在创建磁盘镜像设备时分配的。一个服务最多可以设置 16 个磁盘镜像设备。 |
| 磁盘分区 | 磁盘分区可以为物理磁盘分区也可以是磁盘镜像设备。一个服务最多可以设置 16 个磁盘分区。 |

| 服务属性或资源 | 说明 |
|---|---|
| 检查磁盘 | 指定是否对磁盘进行检查。因为对 SCSI 磁盘和其它磁盘的检查方式不同，所以需要指定磁盘是否是 SCSI 磁盘。还需要指定检查磁盘的 timeout。 |
| 挂载点、文件系统类型和挂载选项 | 如果您的服务要使用某个文件系统，则您必须指定该文件系统的类型、一个挂载点和任何挂载选项。您可以指定的挂载选项应是在 mount.8 manpage 中介绍的标准文件系统挂载选项。如果您的服务使用裸设备，则无需指定挂载信息。<br>ext3等日志文件系统是推荐使用的文件系统。<br>此外，您必须确定是否希望为某个文件系统启用强制卸载（forced unmount）功能。强制卸载功能可使集群服务管理基础设施卸载正在被应用或用户访问的文件系统（就是说，即使文件系统处于“繁忙”状态时也不例外）。 |
| 磁盘分区的所有者、组和访问模式 | 您可以为每一磁盘分区的挂载点或裸设备指定所有者、组和访问模式（例如755）。 |
| 服务启动 timeout | 您可以指定服务启动的timeout,对没有使用磁盘镜像设备的服务，一般服务启动的timeout直接选择默认值就可以了。 |
| 服务停止 timeout | 您可以指定服务停止的timeout,对没有使用磁盘镜像设备的服务，一般服务停止的timeout直接选择默认值就可以了。 |
| 服务停止失败时 reboot | 您可以指定服务停止失败时是否reboot机器。当服务停止失败时，服务的状态将变为error，服务所占用的资源可能并没有释放，选择reboot可以自动释放服务的资源。 |
| 禁用服务政策 | 如果您不希望将某项服务添加到集群后就自动启动它，则可以将其设为禁用状态，直到管理员明确启用该项服务为止。 |

### 7.1.2 创建服务脚本

对于包含某一应用的服务，您必须创建一个包含具体指令的脚本(称为服务脚本或用户脚本)来启动和停止该应用（例如，数据库应用程序）。该脚本将通过一个启动或停止参数来调用，并在服务启动时和停止时运行，且应与系统 V init 目录中出现的脚本相似。

在 GreatTurbo Cluster Server 10 安装光盘的 doc/examples 目录下有用户脚本的例子，名字为 usrapp.sh，您可以参照它来编写自己的用户脚本。另外，在例子中还有一些标准的应用的用户脚本(例如，Oracle 等)，用户只要稍加改动或不用改动即可使用。

在编写用户脚本时，需要注意以下的问题：

1) **脚本的参数必须包含对输入参数 start、stop 的相应处理，在启动部分加入用户所有程序的启动命令，在停止部分加入用户所有程序的停止命令。也就是说脚本的格式必须是规定的格式。请参照例子进行编写。**

2) 请在脚本中加入 log 信息，以便发生故障时准确定位故障原因。

3) 脚本执行成功时要明确返回 0，失败时要明确返回 1。特别是当应用程序有标准的启动/停止脚本时，请确认该脚本在成功时是否返回 0，失败时是否返回 1，如果不是，请不要直接使用应用程序提供的脚本，请手动编写脚本。

4) **对于操作系统默认安装的一些应用的脚本，如 httpd，该脚本的返回值在启动/停止时会判断是否已经启动/停止了，在已经启动/停止时，该脚本返回值是 1，这不满足 GreatTurbo HA 的接口，需要将其改成 0。当然，用户自写的脚本在判断是否已经启动/停止时,也应该满足返回值为 0 的条件。**

5) 如果启动和停止分别有多步操作，停止操作的执行顺序一定要和启动操作的执行顺序相反。

6) 用户脚本要有可执行权限。

用户脚本写好后，请务必通过测试检查用户脚本是否正确。

例如，用户脚本的名字为 usrapp.sh。

1) 启动

首先执行 usrapp.sh　start。

然后 echo $?，判断脚本的返回值是否为 0，0 表示脚本执行成功。

检查应用程序是否正常启动。

如果脚本的返回值为 0 并且应用程序正常启动了，表示用户脚本在启动时是正确的。

2) 停止

首先执行 usrapp.sh　stop。

然后 echo $?，判断脚本的返回值是否为 0，0 表示脚本执行成功。

检查应用程序是否正常停止。

如果脚本的返回值为 0 并且应用程序正常停止了，表示用户脚本在停止时是正确的。

用户服务脚本的一个例子如下所示：

```
#!/bin/sh

# userscript sample

LOG_EMERG=0          # system is unusable
LOG_ALERT=1          # action must be taken immediately
LOG_CRIT=2           # critical conditions
LOG_ERR=3            # error conditions
LOG_WARNING=4        # warning conditions
LOG_NOTICE=5         # normal but significant condition
```

```
LOG_INFO=6                      # informational
LOG_DEBUG=7                     # debug-level messages

script_name=`basename $0`

clulog()
{
        log_level=$1
        log_info=$2
        /opt/cluster/bin/clulog -p $$ -n $script_name -s $log_level "$log_info"
}

case "$1" in
start)
  # start your application, put actual start actions here
    su - oracle -c "dbstart"
    if [ $? -eq 0 ]; then
        # should check application process here
        pmon=`ps -ef | egrep ora_pmon_$ORACLE_SID | grep -v grep`
        if [ "$pmon" = "" ];
        then
        clulog $LOG_ERR " oracle database start failed, process not exist."
        # must return 1 here
            exit 1
        fi
        clulog $LOG_INFO " dbstart succeeded."
    else
        clulog $LOG_ERR " dbstart failed, ret=$?."
        # must return 1 here
```

```
        exit 1
    fi
    su - oracle -c "lsnrctl start"
    if [ $? -eq 0 ]; then
        clulog $LOG_INFO " lsnrctl start succeeded."
        # must return 0 here
        exit 0
    else
        clulog $LOG_ERR " lsnrctl start failed, ret=$?."
        # must return 1 here
        exit 1
    fi
           ;;

stop)
    #stop your application, put actual stop actions here
    su - oracle -c "lsnrctl stop"
    if [ $? -eq 0 ]; then
        clulog $LOG_INFO " lsnrctl stop succeeded."
    else
        clulog $LOG_ERR " lsnrctl stop failed, ret=$?."
               # must return 1 here
        exit 1
    fi
    su - oracle -c "dbshut"
    if [ $? -eq 0 ]; then
        pmon=`ps -ef | egrep ora_pmon_$ORACLE_SID | grep -v grep`
        if [ "$pmon" != "" ];
        then
```

```
            clulog $LOG_ERR " oracle database shut failed, process still exist."
            # must return 1 here
                exit 1
            fi
            clulog $LOG_INFO " dbshut succeeded."
            # must return 0 here
            exit 0
        else
            clulog $LOG_ERR " dbshut failed, ret=$?."
            # must return 1 here
            exit 1
        fi
        ;;
esac
```

### 7.1.3 应用代理 API

GreatTurbo Cluster Server 10 提供了服务检查的功能，可监视集群所支持的单项服务的状态。服务检查不仅包括了对硬件和系统软件的检查，还包括了对特定的应用(例如，数据库服务应用或 HTTP 应用等)的检查。

如果 GreatTurbo Cluster Server 10 发现某一服务出现故障，它将服务从出现故障的集群节点切换到正常运行的集群节点，保持服务的不间断运行。

对服务检查的功能，GreatTurbo Cluster Server 10 提供了灵活的应用代理 API，它是应用代理或服务检查程序与 GreatTurbo Cluster Server 10 服务检查进程之间的接口。借助此 API，您可以为您的服务编写一个定制应用代理。编写定制应用代理的好处在于，它能为您的应用提供更精确的服务检查和更快速的故障切换。

应用代理可以是任何 Linux 可执行的程序，包括 C 程序二进制可执行文件、shell 脚本和 perl 脚本等。根据应用的实际情况，您可以自己编写检测脚本，也可以使用 GreatTurbo Cluster Server 10 自带的位于/opt/cluster/usercheck/目录下的应用代理。

#### 7.1.3.1 检测脚本

在 GreatTurbo Cluster Server 10 安装光盘的 doc/examples 目录下有检测脚本的例子，名字为 usrchk.sh，您可以参照它来编写自己的检测脚本。另外，在例子中还有一些标准的应用的检测脚本(例如，Oracle 等)，用户只要稍加改动或不用改动即可使用。

在编写检测脚本时，需要注意以下的问题：

1) 请在脚本中加入 log 信息，以便发生故障时准确定位故障原因。为减小 log 文件的大小，检查成功时请不要打印 log 信息。

2) 脚本执行成功时返回 0，失败时返回非 0 值。当有多步检查时，请将每步检查失败时的返回值设置为不同的值。

3) 服务检测脚本的名称不要和待检测服务应用程序进程的名字相同。

4) 检测脚本要有可执行权限。

检测脚本写好后，请务必测试检测脚本是否正确。

例如，检测脚本的名字为 usrchk.sh。

1) 应用程序正常运行时

执行检测脚本 usrchk.sh。

然后 echo $?，判断脚本的返回值是否为 0。

如果脚本的返回值为 0，表示检测脚本是正确的。

2) 应用程序运行不正常时

执行检测脚本 usrchk.sh。

然后 echo $?，判断脚本的返回值是否为非 0 值。

如果脚本的返回值为非 0 值，表示检测脚本是正确的。

下面是一个用户检测脚本的例子：

```
#!/bin/sh

# checkscript sample

LOG_EMERG=0                         # system is unusable
LOG_ALERT=1                         # action must be taken immediately
LOG_CRIT=2                          # critical conditions
LOG_ERR=3                           # error conditions
LOG_WARNING=4                       # warning conditions
LOG_NOTICE=5                        # normal but significant condition
LOG_INFO=6                          # informational
LOG_DEBUG=7                         # debug-level messages

script_name=`basename $0`

clulog()
{
        log_level=$1
        log_info=$2
        /opt/cluster/bin/clulog -p $$ -n $script_name -s $log_level "$log_info"
```

```
}

# check applications'process parts, you can add or delete part to check your actual processes

/opt/cluster/usercheck/oracleCheck 172.16.74.127 1521
if [ $? -ne 0 ]
then
        # please add log information here
        clulog $LOG_ERR "oracleCheck failed, ret=$?."
        exit 1
fi

ps -ef | grep ora_pmon_ora10g | grep -v $0 | grep -v grep > /dev/null
if [ $? -ne 0 ]
then
        # please add log information here
        clulog $LOG_ERR "check oracle instance process failed."
        exit 2
fi

ps -ef | grep tnslsnr | grep -v $0 | grep -v grep > /dev/null
if [ $? -ne 0 ]
then
        # please add log information here
        clulog $LOG_ERR " check oracle listener process failed."
        exit 3
fi
```

```
# if all checks aren't wrong, then return 0

exit 0
```

#### 7.1.3.2 GreatTurbo Cluster Server 10 提供的代理

GreatTurbo HA 10 提供了许多典型应用的应用代理。它们包括：

- Sendmail
- Apache
- Oracle
- Samba
- DB2
- DNS
- Informix
- Sybase
- IBM Small Business Suite
- Genic 等

通用应用代理可供那些没有其自身代理的服务使用。它可尝试连接到此类服务的网络端口。如果连接失败，则服务将被视为出现了故障，同时将引发故障切换。

利用这些代理程序，并结合检测脚本，您可以非常自由灵活地定制您的服务检测策略。

### 7.1.4 配置服务的磁盘存储设备

如果您的服务需要使用磁盘存储设备，那么在配置服务之前您必须配置您的磁盘存储设备。

GreatTurbo Cluster Server 10 支持的磁盘存储设备有 2 种：

- 共享磁盘设备
- 磁盘镜像设备

#### 7.1.4.1 配置共享磁盘设备

配置共享磁盘设备的详细方法请参照配置共享磁盘存储器部分的相关说明。

共享磁盘设备配置完成后，您需要使用相关命令(例如，fdisk)对共享磁盘设备进行分区。

在创建您的服务之前，您需要决定您的磁盘设备需要加载文件系统还是使用裸设备。

如果要使用文件系统，您需要利用 mkfs 等命令为您的分区创建文件系统。如欲获得最佳性能，在创建文件系统时应使用 4 KB 的块大小。注意，某些 mkfs 文件系统工具将缺省创建为 1 KB 的块大小，这样可能导致 fsck 的时间长出几倍。对于大型文件系统，使用带有 nocheck 选项的加载命令来绕过检查分区上所有块组的代码。指定 nocheck 选项可显著降低加载大文件系统所需的时间。

如果您的服务使用裸设备，则您可以使用 rawio 文件在引导时间绑定该设备。编辑文件并指定系统启动时您想绑定的裸字符设备和块设备。

#### 7.1.4.2 配置磁盘镜像设备

由于在 Linux kernel2.4 和 2.6 的环境下，drbd 的版本不同，因此配置磁盘镜像设备的方法也不同。

(一) Linux kernel2.4 环境

在 Linux kernel 2.4 环境下，您可以使用 cluadmin 工具来配置磁盘镜像设备。以下举例说明如何使用 cluadmin 来添加磁盘镜像设备。

(1) 在两个节点上分别为磁盘镜像设备准备大小相同的磁盘分区。

(2) 利用 cluadmin 工具配置磁盘镜像设备。

配置磁盘镜像设备的过程如下：

1)执行 nbd add，输入 nbd ID。nbd ID 是镜像设备的编号，必须是 0-15 之间的整数。

```
Cluadmin> nbd add
Currently defined nbds:
Nbd ID(nbd identifier. e.g. 0): 0
```

2)配置节点 A 的 nbd 信息。

```
Nbd configuration on member0:
name(nbd member ip address. e.g. 192.168.0.1): 192.168.1.11
device(block device used by nbd. e.g. /dev/hda14): /dev/hda7
port(port number used by nbd service. e.g. 8787): 8787
deviceNode(nbd block device. e.g. /dev/nb0): /dev/nb0
```

3)配置节点 B 的 nbd 信息。

```
Nbd configuration on member1:
name(nbd member ip address. e.g. 192.168.0.1): 192.168.1.21
device(block device used by nbd. e.g. /dev/hda14): /dev/hda7
port(port number used by nbd service. e.g. 8787): 8787
```

```
deviceNode(nbd block device. e.g. /dev/nb0): /dev/nb0
```

4)确认添加 nbd 设备。

```
Are you sure? (yes/no/?) y
Add nbd device 0
```

**在配置 nbd 设备时需要注意：**

1) **nbd 节点的 IP 必须采用双机直连的 IP。**
2) **同一个 nbd 设备在两节点上的端口号要相同。**
3) **磁盘镜像的设备名必须是/dev/nb#，并且同一个 nbd 设备在两节点上的设备名要相同。**
4) **可以使用 nbd show 显示已添加的 nbd 设备；nbd delete 删除 nbd 设备。**
5) **磁盘镜像设备配置完成后，将生成配置文件/etc/drbd.conf，请不要手动修改这个文件。**

(3) 启动磁盘镜像设备。

在两节点上分别执行命令 /etc/init.d/drbd start。

(4) 初始化磁盘镜像设备。

1)在节点 A 执行 control 0，使节点 A 成为第 0 号磁盘镜像设备的 Primary 节点。(如果为 drbd1，则须执行 control 1,其他依次类推)

```
[root@test1 root]# control 0
Setting 'drbd0' to Primary .. OK
```

```
datadisk: 'drbd0' activated
```

2)查看磁盘镜像设备的状态。

```
 [root@test1 root]# cat /proc/drbd
version: 0.6.8 (api:63/proto:62)

0: cs:Connected st:Primary/Secondary ns:0 nr:0 dw:0 dr:0 pe:0 ua:0
        NEEDS_SYNC
```

3)如果磁盘镜像设备的状态显示为“ NEEDS_SYNC ”，需要同步两节点的磁盘镜像设备。

```
root@test1 root]# datasync -l 0
Option l '0'
…
datadisk: 'drbd0' full synced!
```

4)同步的过程中，可以查看进度。

```
[root@test1 root]# cat /proc/drbd
version: 0.6.8 (api:63/proto:62)

0:  cs:SyncingAll  st:Primary/Secondary  ns:119100  nr:0  dw:0  dr:119108
```

```
pe:253 ua:0
        [==>.................] sync'ed: 14.8% (89611/104391)K
        finish: 0:06min speed: 14,780 (14,780) K/sec
```

5)同步完成后，在 Primary 的节点上创建文件系统。例如，

```
[root@test1 root]# mkfs.ext3 /dev/nb0
```

6)如果需要在磁盘镜像设备上安装应用程序的共享数据，需要执行本步；否则，跳过本步。

mount Primary 节点的磁盘镜像设备，安装应用程序的共享数据到磁盘镜像设备。例如，

```
[root@test1 root]# mount /dev/nb0 /mnt/hda7
…
```

至此，磁盘镜像设备配置完成，然后就可以象使用普通的磁盘一样地使用磁盘镜像设备了。

(二) Linux kernel2.6 环境

在 Linux kernel 2.6 环境下，配置磁盘镜像设备的过程如下：

(1)在两个节点上分别为磁盘镜像设备准备大小相同的磁盘分区。

(2)编辑磁盘镜像设备的配置文件。

在安装 GreatTurbo Cluster Server 10 时，会安装 drbd 的配置文件：/etc/drbd.conf，安装后的 drbd 的配置文件的内容如下所示：

```
#
# /etc/drbd.conf
#

# this is an example of drbd.conf
# please modify the following items according to your real environment.
# - hostname
# - device
# - disk
# - address
# if no particular reasons, you no need to modify other items.

resource drbd0 {
    protocol C;

    startup {
        wfc-timeout    30;
        degr-wfc-timeout 60;
    }

    syncer {
        rate 600M;
        group 0;
    }

    on hostname1 {
        device          /dev/drbd0;
        disk            /dev/hda6;
        address         192.168.0.1:7788;
```

```
        meta-disk    internal;
    }

    on hostname2 {
        device       /dev/drbd0;
        disk         /dev/hda6;
        address      192.168.0.2:7788;
        meta-disk internal;
    }
}
```

对/etc/drbd.conf 需要根据应用的实际环境编辑以下的内容：

- hostname，机器名。为 GreatTurbo HA 节点的机器名。
- device，设备名。为 drbd 设备名，可以选择/dev/drbd0，/dev/drbd1 等，两个节点所对应的 drbd 设备应当一致。
- disk，磁盘分区。为 drbd 设备所对应的物理磁盘分区。
- address，IP 地址和端口。为 drbd 设备通讯所用的网络及端口。Drbd 设备所用的网络最好是直连网络，不同的 drbd 设备所用的端口应当不同。

**注意：**

1) 关于 drbd 配置文件的详细说明请参照 GreatTurbo Cluster Server 10 用户手册。

2) 由于 meta-disk 需要占用 128MB 的磁盘空间，所以 drbd 所用的磁盘分区应当为 128MB 以上。

3) 由于 drbd 采用以太网络进行数据的传输和同步，所以 drbd 通讯所用的网络应当为直连网络，并且最好采用 bonding。

4) 如果要用多个 drbd 设备，请复制整个 resource 部分，并修改 resource 的名字，例如 reource1，然后再相应地对 hostname，device，disk 和 address 进行修改即可。

(3)启动磁盘镜像设备。

在两节点上分别执行命令 /etc/init.d/drbd start。

例如：

```
[root@test1 root]# /etc/init.d/drbd start
Starting DRBD resources:        [ d0 s0 n0 ].
```

(4)查看磁盘镜像设备的状态。

执行命令 cat /proc/drbd 来查看磁盘镜像设备的状态。

如果磁盘镜像设备的数据已经同步，状态显示如下：

```
[root@test1 root]# cat /proc/drbd
version: 0.7.11 (api:77/proto:74)
SVN Revision: 1912M build by root@qa3-127, 2005-08-25 10:20:28
 0: cs:Connected st:Secondary/Secondary ld:Consistent
    ns:0 nr:0 dw:0 dr:0 al:0 bm:92 lo:0 pe:0 ua:0 ap:0
```

如果磁盘镜像设备的数据没有同步，状态显示如下：

```
[root@test1 root]# cat /proc/drbd
version: 0.7.11 (api:77/proto:74)
SVN Revision: 1912M build by root@qa3-127, 2005-08-25 10:20:28
 0: cs:Connected st:Secondary/Secondary ld:Inconsistent
    ns:0 nr:0 dw:0 dr:0 al:0 bm:46 lo:0 pe:0 ua:0 ap:0
```

(5)使其中一个节点成为 Primary 节点。

如果状态显示为“ ld: Consistent ”，请在其中一个节点上执行“ drbdsetup /dev/drbd# primary ”。例如，

```
[root@test1 root]# drbdsetup /dev/drbd0 primary
[root@test1 root]# cat /proc/drbd
version: 0.7.11 (api:77/proto:74)
SVN Revision: 1912M build by root@qa3-127, 2005-08-25 10:20:28
 0: cs:Connected st:Primary/Secondary ld:Consistent
    ns:0 nr:366908 dw:366908 dr:0 al:0 bm:92 lo:0 pe:0 ua:0 ap:0
```

如果状态显示为“ ld: Inconsistent ”,请在其中一个节点上执行“ drbdsetup /dev/drbd# primary --do-what-I-say ”。这时，drbd 会自动进行同步。例如，

```
[root@test1 root]# drbdsetup /dev/drbd0 primary --do-what-I-say
[root@test1 root]# cat /proc/drbd
version: 0.7.12 (api:77/proto:74)
SVN Revision: 1926M build by root@dev3-76, 2005-09-02 10:11:58
 0: cs:SyncSource st:Primary/Secondary ld:Consistent
    ns:152776 nr:0 dw:0 dr:160832 al:0 bm:1118 lo:756 pe:1418 ua:2014 ap:0
        [>...................] sync'ed:    2.5% (5776/5919)M
        finish: 0:02:00 speed: 49,036 (49,036) K/sec
```

(6)创建文件系统。

如果需要使用文件系统，则需要在 Primary 节点上创建文件系统。

```
[root@test1 root]# mkfs.ext3 /dev/drbd0
```

**注意：**

文件系统只需要在 Primary 节点上创建，Secondary 节点会自动进行创建。

(7)如果需要在磁盘镜像设备上安装应用程序的共享数据，需要执行本步操作；否则，跳过本步。

mount Primary 节点的磁盘镜像设备，安装应用程序的共享数据到磁盘镜像设备。例如，

```
[root@test1 root]# mount /dev/drbd0 /opt/oradata
…
```

(8)共享数据安装完成后，需要 umount 磁盘镜像设备。

```
[root@test1 root]# umount /dev/drbd0
```

(9)使 Primary 节点成为 Secondary 节点。

```
[root@test1 root]# drbdsetup /dev/drbd0 secondary
[root@test1 root]# cat /proc/drbd
version: 0.7.11 (api:77/proto:74)
SVN Revision: 1912M build by root@qa3-127, 2005-08-25 10:20:28
 0: cs:Connected st:Secondary/Secondary ld:Consistent
     ns:20348 nr:366908 dw:387256 dr:14468 al:47 bm:139 lo:0 pe:0 ua:0 ap:0
```

至此，磁盘镜像设备配置完成，然后就可以象使用普通的磁盘一样地使用磁盘镜像设备了。

# 7.2 用文本工具 cluadmin 配置和管理服务

Great Cluster Server10 包含两种服务：HA 服务和 LB 服务。cluadmin 可以完成对这两种服务的添加、配置、管理。

## 7.2.1 添加服务

用 cluadmin 工具添加服务的命令是 service add，添加服务的过程如下：

(1) 执行 cluadmin

```
[root@test1 root]# cluadmin
Sat Apr   9 17:48:57 CST 2005
…
```

(2) 执行 service add

GreatTurbo Cluster Server 10 最多只支持 16 个 HA 服务。

```
cluadmin> service add
…
Currently defined services:
```

(3) 输入服务的名字

如果有多个服务，服务的名字不能重复。

```
Service name: svc01
```

(4) 输入服务的类别

选择服务类别是 ha 服务还是 lb 服务。

```
Service type (ha/lb): ha
```

(5) 配置服务的优先节点

如果一个服务有优先节点，当优先节点启动 HA 时或者优先节点的网卡故障恢复时，这个服务将会自动迁移到优先节点上运行。

优先节点缺省值为 None。直接回车表示选择 None，即不选择优先节点。否则输入优先节点的 hostname。

```
Preferred member [None]: test1
```

(6) 配置服务的用户脚本

一个服务最多只能配置一个用户脚本。配置服务的用户脚本时，请输入全路径名。

```
User script (e.g., /usr/foo/script or None) [None]: /etc/init.d/ httpd
```

(7) 配置服务检测脚本

“ Check script ”输入的脚本有两种选择方式，请根据应用选择其中的一种。

第一种方式是使用 GreatTurbo HA 自带的应用 agent，该 agent 位于 /opt/cluster/usercheck 目录下，使用 agent 时，必须按规定格式输入，例如：/opt/cluster/usercheck/httpCheck 172.16.70.100 80。

第二种方式使用自己编写脚本。

配置服务的检测脚本时，必须输入全路径名。

配置服务检测脚本时还需要指定如下的参数：

- “ Check interval ” 是检测服务的时间间隔，默认值是 5 秒，表示每隔 5 秒时间执行一次检测脚本。
- “ Check timeout ” 是检测脚本执行的超时时间，默认值是 30 秒。
- “ Max error count ” 是允许服务连续错误的次数，默认值是 3 次，表示出错 3 次后，服务进行切换。

```
Do you want to add a check script to the service (yes/no/?) [no]: yes
Check Script Information
Check script (e.g., "/opt/cluster/usercheck/httpCheck 172.16.69.200 80" or None) [None]: /opt/cluster/usercheck/httpCheck 172.16.70.100 80
Check interval (in seconds) [None]: 5
Check timeout (in seconds) [None]: 30
Max error count [None]: 3
```

(8) 配置服务的浮动 IP

服务可以绑定浮动 IP，浮动 IP 随着服务而浮动，也就是说这个 IP 所在的节点也就是服务所在的节点。一个服务最多可以配置 16 个浮动 IP。

配置浮动 IP 时必须正确输入如下参数：

- “ Net interface ” 是浮动 IP 绑定的网卡。
- “ Netmask ” 是浮动 IP 的子网掩码。
- “ Broadcast ” 是浮动 IP 的广播地址。

```
Do you want to add floating IP address to the service (yes/no/?) [no]: yes

        IP Address Information

IP address: 172.16.70.100
Net interface [None]: eth0
Netmask (e.g. 255.255.255.0 or None) [None]: 255.255.255.0
```

```
Broadcast (e.g. X.Y.Z.255 or None) [None]: 172.16.71.255
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you (f)inished adding IP addresses: f
```

(9) 配置磁盘设备的信息

GreatTurbo Cluster Server 10 最多可以支持 16 个磁盘设备。

当磁盘设备需要加载文件系统时，必须输入以下的参数：

- “ Device special file ” 是磁盘分区的设备名。对磁盘镜像设备，应为 /dev/nb#。
- “ Check service device ”指定是否检查磁盘设备，默认为 yes。对磁盘镜像设备，请选择 no。
- “ Device is SCSI ”指定磁盘设备是否是 SCSI 设备，默认为 yes，请确认一下您的共享磁盘设备的类型,如果不是 SCSI 设备请选择 no。
- “ Device check timeout ” 是磁盘检测的超时时间，默认为 120 秒。一般情况下选默认值就可以，如果磁盘负载比较大，请适当调大超时时间。
- “ Filesystem type ” 是文件系统的类型。
- “ Mount point ” 是 mount 的位置。
- “ Mount options ” 是 mount 时的选项，一般用 rw, sync。
- “ Forced unmount support ” 是指定 umount 时是否杀死该分区上运行的进程。
- “ Device owner ” 是 mount 时的用户名，一般为 root。
- “ Device group ” 是 mount 时的用户组名，一般为 root。
- “ Device mode ” 是 mount 时的访问权限，一般为 755。

```
Do you want to add a disk device to the service (yes/no/?) [no]: yes
```

```
          Disk Device Information

Device special file (e.g., /dev/sda1): /dev/sdb0
Check service device (yes/no/?) [yes]:
Device is SCSI (yes/no/?) [yes]:
Device check timeout (in seconds) [120]:
Filesystem type (e.g., ext2, reiserfs, ext3 or None): ext3
Mount point (e.g., /usr/mnt/service1) [None]: /mnt/userData
Mount options (e.g., rw,nosuid): rw
Forced unmount support (yes/no/?) [no]: yes
Device owner (e.g., root): root
Device group (e.g., root): root
Device mode (e.g., 755): 755
Do you want to (a)dd, (m)odify, (d)elete or (s)how devices, or are you (f)inished
adding devices: f
```

一次只能输入一个分区设备的信息，如果有多个分区设备，可以在结束提示时，不选择 f，而选择 a 继续增加分区设备。

当服务所用的磁盘设备是裸设备时，只需输入磁盘分区的设备名称，及指定磁盘检测的相关信息，其他信息必须全部按回车键略过。

```
Do you want to add a disk device to the service (yes/no/?) [no]: yes

          Disk Device Information

Device special file (e.g., /dev/sda1): /dev/sdc
Check service device (yes/no/?) [yes]:
Device is SCSI (yes/no/?) [yes]:
Device check timeout (in seconds) [120]:
Filesystem type (e.g., ext2, reiserfs, ext3 or None):
```

```
Mount point (e.g., /usr/mnt/service1) [None]:
Mount options (e.g., rw,nosuid):
Forced unmount support (yes/no/?) [no]:
Device owner (e.g., root):
Device group (e.g., root):
Device mode (e.g., 755):
Do you want to (a)dd, (m)odify, (d)elete or (s)how devices, or are you (f)inished
adding devices: f
```

同样一次只能输入一个裸设备的名字，如果有多个裸设备，可以在结束提示时，不选择 f，而选择 a 继续增加分区设备。

(10) 设置启动服务的超时时间

对于不使用磁盘镜像设备的服务，建议此处直接回车，以选择服务默认启动超时时间 3600 秒。

对于使用磁盘镜像设备的服务，请按照以下的方法估算服务启动时间。

启动服务的时间 ＝ 启动 nbd 的时间 ＋mount 设备的时间 ＋ 启动 IP 的时间 ＋ 启动用户脚本的时间

启动服务的超时时间应当大于启动服务的时间，各个阶段的时间可以估算来设置。

由于启动 nbd 时，可能会进行两节点间全盘数据的同步，因此比较花费时间。启动 nbd 的时间可以按如下的方法进行估算：

启动 nbd 的时间 ＝ 磁盘分区大小/nbd 网络带宽 ＋ 余量

当配置有多个 nbd 设备时，要分别计算。

例如：系统配置了 2 个 nbd 设备，分区大小分别为 100MB，500MB，使用同一个网络接口，网络带宽为 100Mbps。

启动 nbd 的时间 = (100*8/100 + 500*8/100)*2 + 100

= 196

考虑到 mount 设备的时间、启动 IP 的时间、启动用户脚本的时间，启动服务的超时时间可以设置为 300 秒。

```
Service start timeout (in seconds) [None]:
```

(11) 设置停止服务的超时时间

对于不使用磁盘镜像设备的服务，建议此处直接回车，以选择服务默认停止超时时间 3600 秒。

对于使用磁盘镜像设备的服务，请按照以下的方法估算服务停止时间。

停止服务的时间 = 停止用户脚本的时间 + 停止 IP 的时间 + umount 设备的时间 + 停止 nbd 的时间

停止服务的超时时间应当大于停止服务的时间，各个阶段的时间可以估算来设置。

由于停止 nbd 时，可能会进行两节点间全盘数据的同步，因此比较花费时间。

停止服务的超时时间的估算方法可以参考启动服务的超时时间的估算方法。

```
Service stop timeout (in seconds) [None]:
```

(12) 设置停止服务失败时是否 reboot 机器

如果设为“ yes ”，当服务停止失败时，为了释放服务的资源，将自动 reboot 机器。

如果设为 “ no ” ，当服务停止失败时，不会自动 reboot 机器，需要用户手动干预。

```
Reboot system if stop the service failed (yes/no/?) [yes]: yes
```

(13) 设置是否 disable 服务

如果设为“ yes ”，服务将不会立即被启动，只能以后由用户手动启动。如果设为“ no ”，服务将立即被启动。一般选择 no，表示立即启动服务。

```
Disable service (yes/no/?) [no]: no
```

(14)确认添加该服务

如果确认配置正确，请选择 yes 添加以上配置的服务。

```
Add svc01 service as shown? (yes/no/?) yes
Added svc01.
```

至此，我们成功地添加了一个 GreatTurbo Cluster Server 10 的服务。如果您还想再添加服务，请重复上面的步骤。

### 7.2.2 显示服务配置

如果要显示服务的配置信息，请执行 cluadmin 的 service show config 命令。

```
cluadmin> service show config
  0) svc01
  c) cancel

Choose service: 0
service name: svc01
```

```
disabled: yes
preferred node: test1
user script: /etc/init.d/httpd
check script: /opt/cluster/usercheck/httpCheck 172.16.69.201 80
   check interval: 5
   check timeout: 30
   max error count: 3
IP address 0: 172.16.69.100
   net interface 0: eth1
   netmask 0: 255.255.252.0
   broadcast 0: 172.16.71.255
Nbd device 0: 0
device 0: /dev/nb0
   check device, device 0: no
   mount point, device 0: /mnt/hda8
   mount fstype, device 0: ext3
   mount options, device 0: rw
   force unmount, device 0: yes
   owner, device 0: root
   group, device 0: root
   mode, device 0: 755
start timeout: 300
stop timeout: 300
reboot system if stop service failed: yes
cluadmin>
```

如果您知道服务名称，那么您可以指定服务显示命令 service show config service_name。

### 7.2.3 修改服务

如果需要修改服务的配置，请执行 cluadmin 的 service modify 命令。

首先要求选择需要修改的服务，请输入列表中的服务的编号。

如果服务正在运行，将要求先停止服务，请输入 yes。

然后所有的操作都和添加服务相同。

```
Cluadmin> service modify


  You will be prompted for information about the service.


  Enter a question mark (?) at a prompt to obtain help.


  Enter a colon (:) and a single-character command at a prompt to do
  one of the following:


  c - Cancel and return to the top-level cluadmin command
  r - Restart to the initial prompt while keeping previous responses
  p - Proceed with the next prompt



  0) svc01
  c) cancel

Choose service to modify: 0
Modifying: svc01
Service is not disabled.   Disable it? (yes/no/?) yes
```

### 7.2.4 删除服务

如果要删除服务，请执行 cluadmin 的 service delete 命令。

然后输入列表中的服务的编号即可。

```
cluadmin> service delete
  0) svc01
  c) cancel

Choose service to delete: 0
Deleting svc01, are you sure? (yes/no/?): yes
Service svc01 disabled
Svc01 deleted.
```

### 7.2.5 禁用服务

您可以禁用某一正在运行的服务，来使其不可用。一旦服务被禁用后，下次机器启动时将不会自动启动服务。如果要使其重新可用，您必须手动启用它。

如果需要禁用服务，请执行 cluadmin 的 service disable 命令。

首先选择要禁用的服务，请输入列表中的服务的编号。

然后要求确认，请输入 yes。

```
cluadmin> service disable
  0) svc01
  c) cancel

Choose service to disable: 0
Are you sure? (yes/no/?) yes
```

```
Disabling svc01. Service   disabled.
```

您也可以禁用某一处于错误状态的服务,这可以使服务从错误状态下恢复正常。更多详情，请参见处理错误状态下的服务的相关章节。

### 7.2.6 启用服务

如果在添加服务时没有启动服务，或者服务被禁用时，需要手动启用服务，请执行 cluadmin 的 service enable 命令。

首先要求选择需要启动的服务，请输入列表中的服务的编号。

然后要求选择在哪个节点上启动服务，请输入节点的编号。

```
cluadmin> service enable
  0) svc01
  c) cancel

Choose service to enable: 0
Are you sure? (yes/no/?) yes
  0) test1
  1) test2
  c) cancel

Choose member: 0
Enabling svc01 on member test1. Service   enabled.
```

### 7.2.7 切换服务

除了提供自动服务故障切换功能外，GreatTurbo Cluster Server 10 还可帮助您明确地停止一个节点上运行的服务，然后在另一个集群节点上启动它。服务切换功能可使管理员在一个集群节点上执行维护任务的同时，保持应用和数据的高可用性。

如果需要把在一个节点上运行的服务切换到另一个节点上，请执行 cluadmin 的 service relocate 命令。

首先选择要切换的服务，请输入列表中的服务的编号。

然后要求确认，请输入 yes。

```
cluadmin> service relocate
  0) svc01
  c) cancel

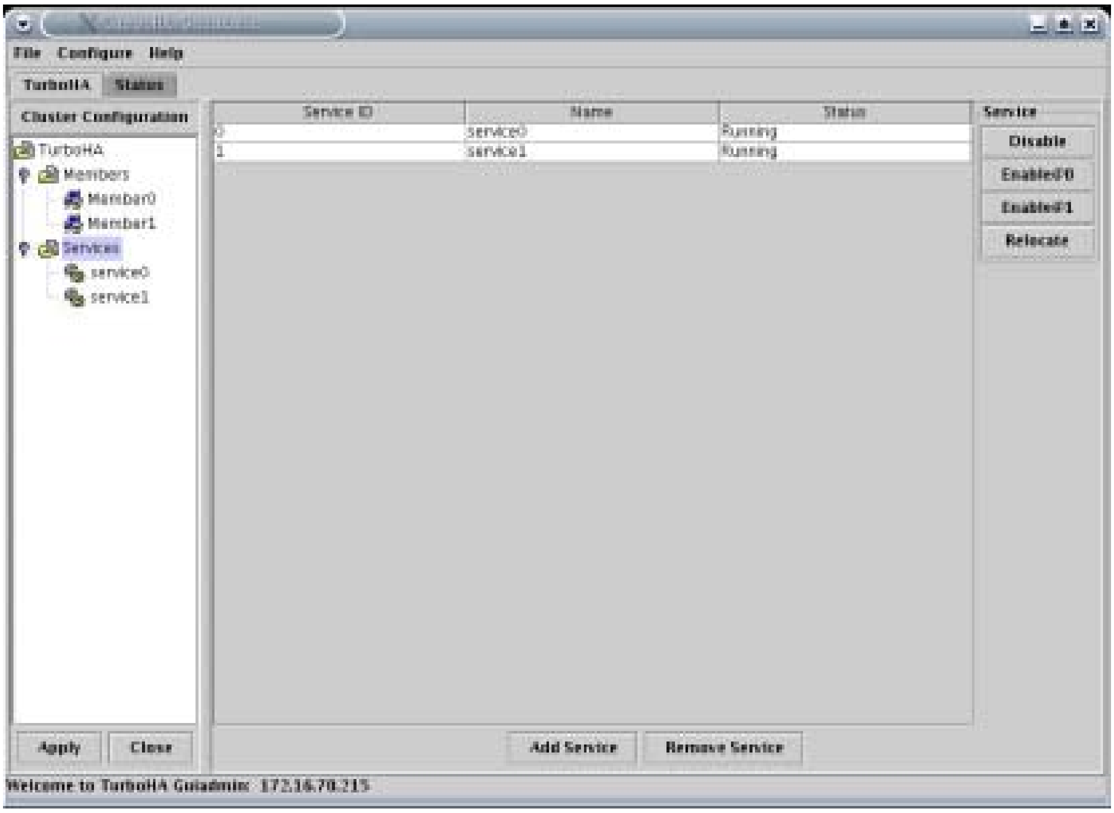
Choose service to relocate: 0
Are you sure? (yes/no/?) yes
Relocating svc01. Service   relocated.
```

## 7.3 用图形工具 guiadmin 配置和管理服务

Great Cluster Server10 包含两种服务：HA 服务和 LB 服务。guiadmin 的服务管理功能只针对 HA 服务，不能完成对 LB 服务的添加、配置、管理。

节点树中的 Services 项用于添加、修改、删除服务和控制服务的状态。请注意：目前图形工具 guiadmin 只能对 HA 服务进行管理。

通过该项的工作窗口，您可以查看所有已配置服务的当前状态，可以添加、删除服务，并可以启动、停止或者切换已配置服务。




下表描述了该窗口的各项特性：

| 字段 | 说明 |
|---|---|
| Service ID | 显示服务的 ID 号。 |
| Name | 显示服务的名称 |
| Status | 显示服务的运行状态。 |

下表描述了可以对服务进行的操作：

| 按钮名 | 说明 |
|---|---|
| Add Service | 增加服务，需要按 “ apply ” 键来确认。 |
| Remove Service | 删除选中的服务，只有被 disable 的服务才可以被删除。该操作直接提交到服务器，不需要按 “ apply ”  健来确认。 |

| 按钮名 | 说明 |
|---|---|
| Disable | 禁用选定的服务。 |
| Enable@0 | 在第 0 个节点上启用所选的服务。 |
| Enable@1 | 在第 1 个节点上启用所选的服务。 |
| Relocate | 使服务在当前节点和另外一个节点之间切换。 |

### 7.3.1 添加服务

点击 Services 窗口中的 Add Service 按钮，如下所示的对话框会弹出，在这个对话框中用户需要输入新服务的名称。

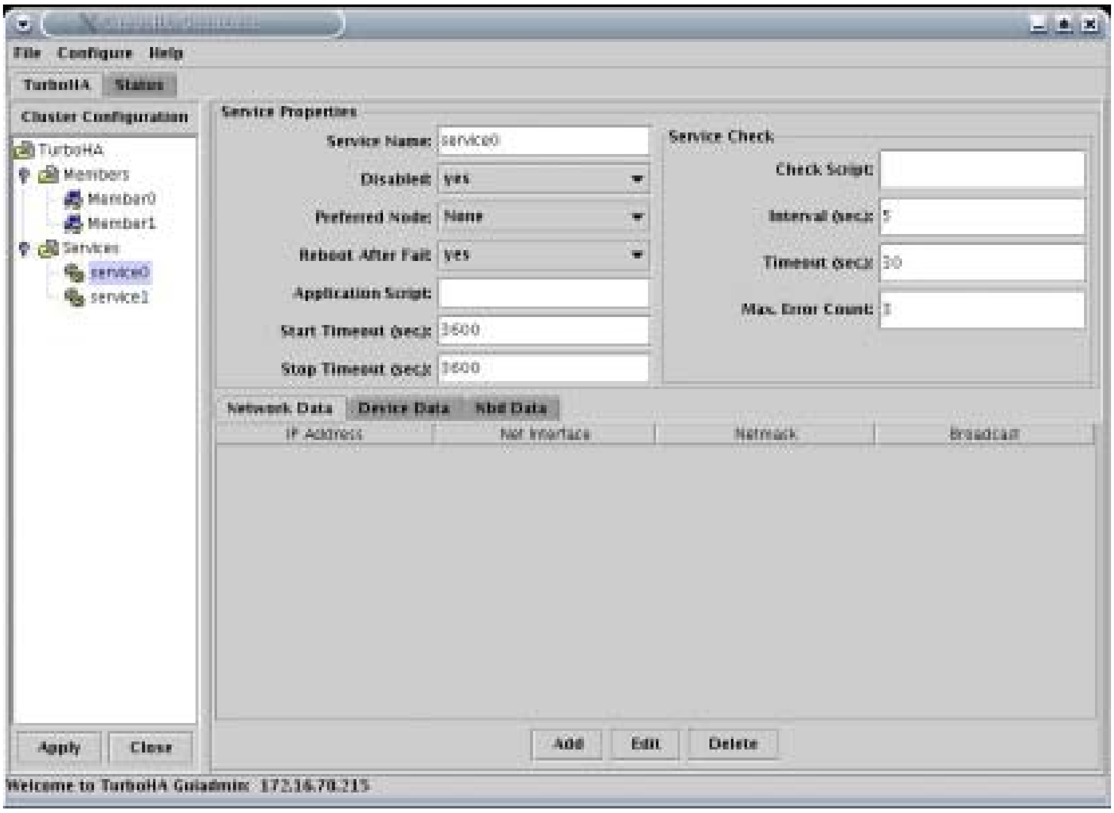



当用户输入完服务名称后，新的服务就会出现在节点树中。 然后按照如下顺序操作，就可以完整的添加一个服务。

1：点击节点树中新建项。

2：按照要求在主页面各字段中填入正确的信息。

3：点击 “ Apply ” 按钮让数据传送到 Server，使用户修改生效。



下表描述了添加服务过程中需要配置的字段的详细说明：

| 字段 | 说明 |
|---|---|
| Service Name | 设置服务名称。 |
| Disabled | 将服务的默认状态设为禁用。默认值是 no，即添加完毕启动服务。 |
| Preferred Node | 设置运行该服务的首选节点。 |
| Reboot After Fail | 服务添加失败后是否重新启动机器。默认为 yes |
| Application Script | 控制服务启动和停止的脚本。 |
| Start Timeout | 设置脚本启动等待时间，默认为 3600 秒。此项值的范围是[60-65535]秒。 |
| Stop Timeout | 设置脚本停止等待时间，默认为 3600 秒。此项值的范围是[60-65535]秒。 |
| Service Check Data | |

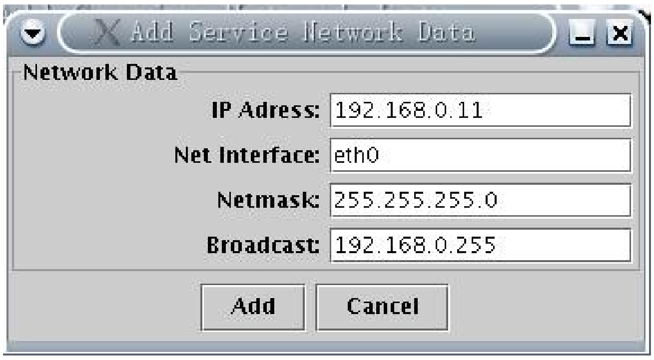
| 字段 | 说明 |
| --- | --- |
| Check Script | 检查服务状态的脚本。 |
| Interval | 服务检查的时间间隔,以秒为单位。默认为 5 秒。此项值的范围是[3-60]秒。 |
| Timeout | 服务状态检查响应的最长等待时间，超时则记录为发生错误。默认为 30 秒。此项值的范围是[30-3600]秒。 |
| Max Error Count | 服务重新定位之前容许的最大错误次数。默认为 3 次。此项值的范围是[1-60]。 |
| Service Network Data | |
| IP Address | 服务对应的浮动 IP 地址。所有服务的此项不能够重复。 |
| Net Interface | 服务所在的 IP 地址所绑定的网卡 |
| Netmask | 网络掩码地址。 |
| Broadcast | 广播地址。如果 ip 地址和网络掩码正确输入的话，此项会自动计算。 |
| Service Device Data | |
| Device File | 存储设备的设备文件名。所有服务的此项不能够重复。 |
| FS Type | 文件系统类型。如果是存储设备是裸设备则不需要输入此项。 |
| Owner | 设备文件所有者。 |
| Group | 设备文件组。 |
| Mode | 设备的八进制权限值。 |
| Mount Point | 存储设备的 mount 点，所有服务的此项不能够重复。如果是裸设备则不需要输入此项。 |

| 字段 | 说明 |
| --- | --- |
| Mount Options | 设定设备的特定加载选项。 |
| Force Unmount | 如果设为 yes，设备将在服务器禁用后强行卸载。 |
| Check Device | 是否检查存储设备。如果此项选择为 no，则以下两项不能够编辑。 |
| SCSI | 存储设备是否为 SCSI 设备。 |
| Check Timeout | 检查设备所需要的延时时间。此项值的范围是[30-3600] |
| Service NBD Data | |
| Nbd ID | 镜像设备的编号，必须是 0-15 之间的整数。 |
| Name0 | 节点 1 上的 Nbd 成员的 ip 地址。例如：192.168.0.1。 |
| Device0 | 节点 1 上被 NBD 设备使用的块设备。例如：/dev/hda14 |
| Port0 | 节点 1 上被 nbd 设备使用的端口号。例如：8787 |
| Device Node0 | 节点 1 上的 Nbd 块设备。例如：/dev/nb0。 |
| Name1 | 节点 2 上的 Nbd 成员的 ip 地址。例如：192.168.0.1。 |
| Device1 | 节点 2 上被 NBD 设备使用的块设备。例如：/dev/hda14 |
| Port1 | 节点 2 上被 nbd 设备使用的端口号。例如：8787 |
| Device Node1 | 节点 2 上的 Nbd 块设备。例如：/dev/nb0。 |

#### 7.3.1.1 添网络数据(Service Network Data)

设置 Network Data 就指定了一个服务对外的 ip 地址及相关一些参数。在节点树中选择所要配置的服务节点，然后在主页面中作如下操作。

1：点击“ Network Data ”标签，然后点击“ Add ”按钮，会弹出如下的对话框；




2：在对话框中按照要求正确填入所需选项；

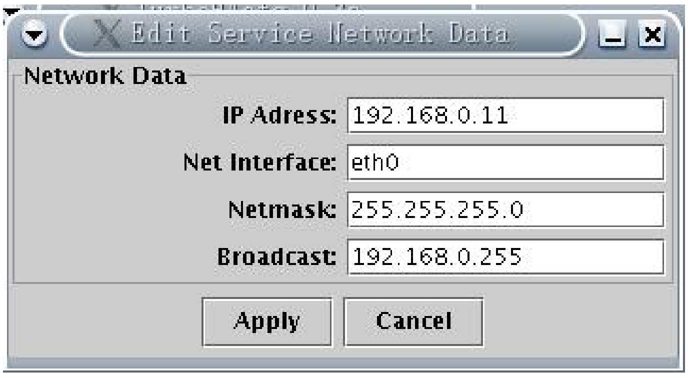
3：点击“ Add ”键，即为该服务添加了一个浮动 ip。

以上操作并没有把数据提交给服务器端，用户必须在修改该服务的信息后点击节点树下的“ Apply ”键，才能够使数据提交到服务器。

#### 7.3.1.2 修改网络数据(Service Network Data)

在节点树中选择所要配置的服务节点，然后在主页面中作如下操作。

1：首先点击“ Network Data ”标签，然后点击“ Edit ”按钮，会弹出如下对话框；




2：在对话框中按照要求修改所需选项；

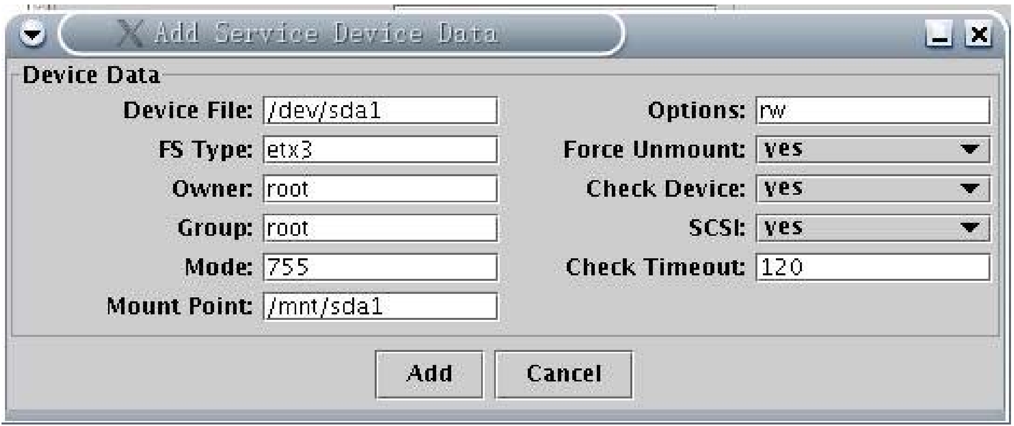
3：点击此页面上的“ Apply ”键，即完成了对网络数据的修改。

以上操作并没有把数据提交给服务器端，用户必须在修改该服务的信息后点击节点树下的“ Apply ”键，才能够使数据提交到服务器。

#### 7.3.1.3 添加服务的设备数据

用户可以在服务中使用磁盘设备，这就需要用户在增加服务的同时对服务的设备进行设置。在节点树中选择所要配置的服务节点，然后在主页面中作如下操作。

1：点击“ Device Data ”标签，然后点击 “ Add ”按钮，会弹出如下所示的对话框；

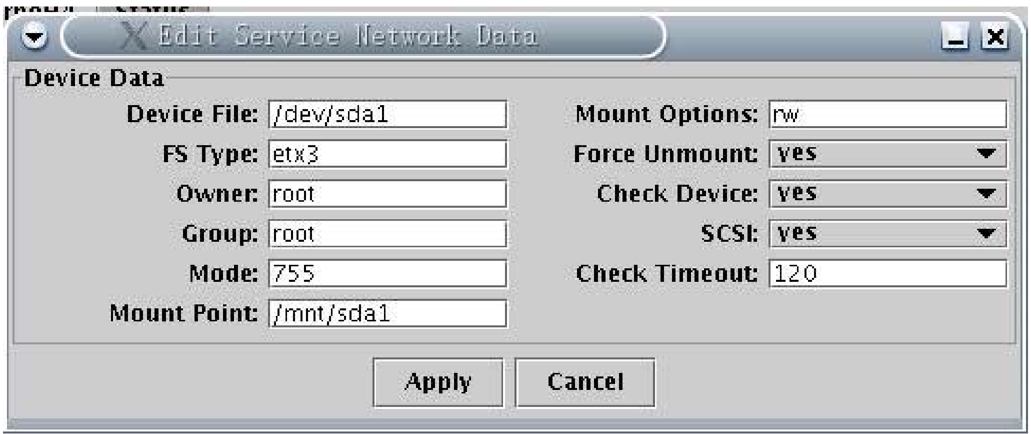



2：在对话框中按照要求正确填入所需选项；

3：点击“ Add ”键，即在服务中添加了一个设备。

#### 7.3.1.4 修改服务的设备数据

1：点击“ Device Data ”标签，然后点击页面最下方的“ Edit ”按钮，会弹出如下的对话框；




2：在对话框中按照要求修改所需选项；

3：点击“ Apply ”键，即完成了对设备数据的修改。

#### 7.3.1.5 查看服务的 Nbd 设备数据

用户可以在服务中使用镜像设备，这是为了满足既想使用磁盘设备，又不具备磁盘阵列条件的用户设计的。用户需要使用 cluadmin 工具来配置服务的 drbd 设备，guiadmin 可以查看服务的 drbd 设备信息。

### 7.3.2 显示服务配置

如果用户希望查看已有服务的配置信息，根据服务名点击节点树上的相应节点，该服务的配置信息将显示在右侧的主页面中。用户可以一目了然的看到服务的全部信息。

### 7.3.3 修改服务

如果用户希望修改已经存在的服务，那么可以按照如下操作进行。

1：点击节点树中想要修改的服务。

2：按照要求正确修改服务信息。

3：点击 Apply 键让数据生效。

### 7.3.4 删除服务

在 Services 窗口选择想要删除的服务，点击“ Remove Servcie ”按钮就完成了删除服务的操作。

要删除一个服务，首先需要把那个服务 disable。不能够删除一个正在运行的服务。

删除服务的操作直接提交到服务器生效，不需要用户点击“ Apply ”键。

### 7.3.5 禁用服务

在 Services 窗口选择想要 disable 的服务，点击“ Disable ”按钮来使服务实效。

一个服务处于 “ running ” 或者 “ error ” 状态才能够 disable。

Disable 服务的操作直接提交到服务器生效，不需要用户点击 “ Apply ” 按钮。

### 7.3.6 启用服务

在 Services 窗口选择想要启动的服务，点击 “ Enable@0 ” 或者 “ Enable@1 ” 按钮来启动服务。“ Enable@0 ” 表示在 member0 上启动，也就是第一个节点；反之就表示在 member1 上启动。

一个服务只有处于 “ not running ” 状态才能够启动，并且只能够在一个节点启动。

启动服务的操作直接提交到服务器，不需要用户点击 “ Apply ” 按钮。

### 7.3.7 切换服务

在 Services 窗口选择想要切换服务，点击 “ Relocate ” 按钮在不同的节点间切换服务。如果服务在第一个节点启动，那么点击 “ Relocate ” 以后，服务将在第一个节点停止运行，在另外一个节点启动。

一个服务只有处于 “ running ” 状态才能够切换。

切换服务的操作直接提交到服务器，不需要用户点击 “ Apply ” 按钮。

## 7.4 配置典型应用的服务

### 7.4.1 配置 Oracle 服务

数据库服务可为数据库应用提供高可用性的数据。之后，该应用即可为数据库客户机系统（比如 web 服务器等）提供网络访问能力。如果服务发生故障切换，则应用将通过新的集群节点访问共享数据库数据。可通过网络访问的数据库通常都指定有一个 IP 地址，它将同服务一起进行故障切换，以保持客户机的透明访问。

本节针对 Oracle 数据库提供了一个设置集群服务的例子。

举例内容如下：

- 服务中包含一个供 Oracle 客户机使用的 IP 地址。
- 服务中拥有两个已加载的文件系统：一个用于 Oracle 软件（/u01），另一个用于 Oracle 数据库（/u02），这两个文件系统均需在添加服务之前设置。
- 添加服务之前，在两个集群系统上均创建了一个带有 oracle 名称的 Oracle 管理账户。

下面是一个用于启动和停止 Oracle 服务的脚本例子。

```
#!/bin/sh
#
# oraShell -- user start|stop script for oracle
#

LOG_EMERG=0        # system is unusable
LOG_ALERT=1        # action must be taken immediately
LOG_CRIT=2         # critical conditions
LOG_ERR=3          # error conditions
```

```
LOG_WARNING=4     # warning conditions
LOG_NOTICE=5      # normal but significant condition
LOG_INFO=6        # informational
LOG_DEBUG=7       # debug-level messages

script_name=`basename $0`

clulog()
{
        log_level=$1
        log_info=$2
        /opt/cluster/bin/clulog -p $$ -n $script_name -s $log_level "$log_info"
}

case $1 in
start)
    dbstart_output=$(su - oracle -c "dbstart")
    if [ $? -eq 0 ]; then
        echo ${dbstart_output} | egrep "Database \"${ORACLE_SID}\" warm
started" >/dev/null 2>&1
        if [ $? -ne 0 ]; then
            clulog $LOG_ERR "oraShell: oracle database start failed, database
\"${ORACLE_SID}\" not started."
            exit 1
        fi
        pmon=`ps -ef | egrep ora_pmon_$ORACLE_SID | grep -v grep`
        if [ "$pmon" = "" ];
        then
            clulog $LOG_ERR "oraShell: oracle database start failed, process not
```

```
exist."
                exit 1
            fi
            clulog $LOG_INFO "oraShell: dbstart succeeded."
        else
            clulog $LOG_ERR "oraShell: dbstart failed, ret=$?."
            exit 1
        fi
        su - oracle -c "lsnrctl start"
        if [ $? -eq 0 ]; then
            clulog $LOG_INFO "oraShell: lsnrctl start succeeded."
            exit 0
        else
            clulog $LOG_ERR "oraShell: lsnrctl start failed, ret=$?."
            exit 1
        fi

        ;;

stop)
        su - oracle -c "lsnrctl stop"
        if [ $? -eq 0 ]; then
            clulog $LOG_INFO "oraShell: lsnrctl stop succeeded."
        else
            clulog $LOG_ERR "oraShell: lsnrctl stop failed, ret=$?."
            exit 1
        fi
        dbshut_output=$(su - oracle -c "dbshut")
        if [ $? -eq 0 ]; then
```

```
            echo ${dbshut_output} | egrep "Database \"${ORACLE_SID}\" shut down" >/dev/null 2>&1
            if [ $? -ne 0 ]; then
                clulog $LOG_ERR "oraShell: oracle database shut failed, database \"${ORACLE_SID}\" not shut down."
                exit 1
            fi
            pmon=`ps -ef | egrep ora_pmon_$ORACLE_SID | grep -v grep`
            if [ "$pmon" != "" ];
            then
                clulog $LOG_ERR "oraShell: oracle database shut failed, process still exist."
                exit 1
            fi
            clulog $LOG_INFO "oraShell: dbshut succeeded."
            exit 0
        else
            clulog $LOG_ERR "oraShell: dbshut failed, ret=$?."
            exit 1
        fi

        ;;

esac
```

下面是一个用于检查 Oracle 服务的脚本例子。

```
#!/bin/sh
#
# oraCheck -- check script for oracle
#

LOG_EMERG=0                 # system is unusable
LOG_ALERT=1                 # action must be taken immediately
LOG_CRIT=2                  # critical conditions
LOG_ERR=3                   # error conditions
LOG_WARNING=4               # warning conditions
LOG_NOTICE=5                # normal but significant condition
LOG_INFO=6                  # informational
LOG_DEBUG=7                 # debug-level messages

script_name=`basename $0`

clulog()
{
        log_level=$1
        log_info=$2
        /opt/cluster/bin/clulog -p $$ -n $script_name -s $log_level "$log_info"
}

# please change the IP to the floating IP of your service
/opt/cluster/usercheck/oracleCheck 172.16.74.127 1521
if [ $? -ne 0 ]
then
        clulog $LOG_ERR "oraCheck: oracleCheck failed, ret=$?."
        exit 1
```

```
fi


ps -ef | grep ora_pmon | grep -v $0 | grep -v grep >/dev/null 2>&1
if [ $? -ne 0 ]
then
        culog $LOG_ERR "oraCheck: check oracle instance process failed."
        exit 2
fi


ps -ef | grep tnslsnr | grep -v $0 | grep -v grep >/dev/null 2>&1
if [ $? -ne 0 ]
then
        culog $LOG_ERR "oraCheck: check oracle listener process failed."
        exit 3
fi


exit 0
```

以下举例说明了如何使用 cluadmin 来添加 Oracle 服务 。

```
cluadmin> service add


  The user interface will prompt you for information about the service.
    Not all information is required for all services.


  Enter a question mark (?) at a prompt to obtain help.
```

```
  Enter a colon (:) and a single-character command at a prompt to do
  one of the following:

  c - Cancel and return to the top-level cluadmin command
  r - Restart to the initial prompt while keeping previous responses
  p - Proceed with the next prompt

Currently defined services:

Service name：oracle
Preferred member [None]：test1
User script（e.g., /usr/foo/script or None）[None]：/home/oracle/oraShell
Do you want to add a check script to the service (yes/no/?) [no]: yes

        Check Script Information

Check script (e.g., "/opt/cluster/usercheck/httpCheck 172.16.69.200 80" or None)
[None]: /home/oracle/oraCheck
Check interval (in seconds) [None]: 5
Check timeout (in seconds) [None]: 30
Max error count [None]: 8
Do you want to (m)odify, (d)elete or (s)how the check script, or are you (f)inished
adding check script: f
Do you want to add floating IP address to the service (yes/no/?) [no]: yes

        IP Address Information

IP address：  10.1.16.132
Net interface [None]: eth0
```

```
Netmask（e.g. 255.255.255.0 or None）[None]： 255.255.255.0
Broadcast（e.g. X.Y.Z.255 or None）[None]： 10.1.16.255
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you (f)inished adding IP addresses: f
Do you want to add a nbd mirror disk device to the service (yes/no/?) [no]: no
Do you want to add a disk device to the service（yes/no/?）：yes

Disk Device Information

Device special file（e.g., /dev/sda1）：/dev/sda1
Check service device (yes/no/?) [yes]: yes
Device is SCSI (yes/no/?) [yes]: yes
Device check timeout (in seconds) [120]:
Filesystem type（e.g., ext2, reiserfs, ext3 or None）：ext3
Mount point（e.g., /usr/mnt/service1 or None）[None]：/u01
Mount options（e.g., rw, nosuid）：[Return]
Forced unmount support（yes/no/?）[no]：yes
Device owner（e.g., root）：root
Device group（e.g., root）：root
Device mode（e.g., 755）： 755
Do you want to (a)dd, (m)odify, (d)elete or (s)how devices, or are you (f)inished adding devices: a
Device special file（e.g., /dev/sda1）：/dev/sda2
Check service device (yes/no/?) [yes]:
Device is SCSI (yes/no/?) [yes]:
Device check timeout (in seconds) [120]:
Filesystem type（e.g., ext2, reiserfs, ext3 or None）：ext3
Mount point（e.g., /usr/mnt/service1 or None）[None]：/u02
Mount options（e.g., rw, nosuid）：[Return]
```

```
Forced unmount support（yes/no/?）[no]：yes
Device owner（e.g., root）：root
Device group（e.g., root）：root
Device mode（e.g., 755）： 755
Do you want to (a)dd, (m)odify, (d)elete or (s)how devices, or are you (f)inished adding devices: f
Service start timeout (in seconds) [None]:
Service stop timeout (in seconds) [None]:
Reboot system if stop the service failed (yes/no/?) [yes]:
Disable service（yes/no/?）[no]：no
service name: oracle
disabled：no
preferred node：test1
user script：/home/oracle/oraShell
check script: /home/oracle/oraCheck
   check interval: 5
   check timeout: 30
   max error count: 8
IP address 0： 10.1.16.132
   net interface 0: eth0
netmask 0： 255.255.255.0
broadcast 0： 10.1.16.255
device 0：/dev/sda1
   check device, device 0: yes
   SCSI device, device 0: yes
   check timeout, device 0: 120
mount point, device 0：/u01
mount fstype, device 0：ext3
force unmount, device 0：yes
```

```
owner, device 0: root
group, device 0: root
mode, device 0: 755
device 1：/dev/sda2
   check device, device 1: yes
   SCSI device, device 1: yes
   check timeout, device 1: 120
mount point, device 1：/u02
mount fstype, device 1：ext3
force unmount, device 1：yes
owner, device 0: root
group, device 0: root
mode, device 0: 755
start timeout: None
stop timeout: None
reboot system if stop service failed: yes
Add oracle service as shown? (yes/no/?) yes
   0) test1
   1) test2
   c) cancel

Choose member to start service on: 0
Added oracle.
cluadmin>
```

### 7.4.2 配置 MySQL 服务

数据库服务可为数据库应用提供高可用性的数据。之后，该应用即可为数据库客户机系统（比如 web 服务器等）提供网络访问能力。如果服务发生故障切换，则应用将通过新的集群节点访问共享数据库数据。可通过网络访问的数据库通常都指定有一个 IP 地址，它将同服务一起进行故障切换，以保持客户机的透明访问。

您可以在集群中设置一项 MySQL 数据库服务。注意，MySQL 不提供完整的事务处理语义；因此，它可能不适合于更新密集型应用。

下面是一个 MySQL 数据库服务举例：

- MySQL 服务器和数据库例程均存储在某一文件系统上（位于共享存储上的某一磁盘分区）。这样可使数据库数据及其运行时状态信息（需要进行故障切换）同时被两个集群系统访问。在该举例中，文件系统被作为/var/mysql 加载，使用共享磁盘分区/dev/sda1。
- IP 地址结合 MySQL 数据库可支持数据库服务客户机发起的网络访问。该 IP 地址将以服务故障切换的方式在集群成员中自动进行移植。在下面的举例中，IP 地址为 10.1.16.12。
- 用于启动和停止 MySQL 数据库的脚本是标准的系统 V init 脚本，该脚本已经采用配置参数进行了修改，可与装有数据库的文件系统相匹配。
- 在缺省情况下，连接到某台 MySQL 服务器的客户机在连续 8 小时处于休止状态后，将出现超时现象。您可以在启动 mysqld 时通过设置 wait_timeout 变量来修改这一连接限制。

如果想查看某台 MySQL 服务器是否已经超时，则调用 mysqladmin 版本命令并检查其正常运行时间。然后再次调用查询，自动重新连接到服务器。

根据 Linux 发行版的不同，下列其中一条信息可能表明某台 MySQL 服务器超时：

CR_SERVER_GONE_ERROR

CR_SERVER_LOST

下面是一个用于启动和停止 MySQL 服务的脚本例子。

```
#!/bin/sh
# Copyright Abandoned 1996 TCX DataKonsult AB & Monty Program KB & Detron
HB
# This file is public domain and comes with NO WARRANTY of any kind

# Mysql daemon start/stop script.

# Usually this is put in /etc/init.d（at least on machines SYSV R4
# based systems）and linked to /etc/rc3.d/S99mysql.When this is done
# the mysql server will be started when the machine is started.

# Comments to support chkconfig on RedHat Linux
# chkconfig： 2345 90 90
# description：A very fast and reliable SQL database engine.

PATH=/sbin：/usr/sbin：/bin：/usr/bin
basedir=/var/mysql
bindir=/var/mysql/bin
datadir=/var/mysql/var
pid_file=/var/mysql/var/mysqld.pid
mysql_daemon_user=root   # Run mysqld as this user.
export PATH

mode=$1
```

```
if test -w /                    # determine if we should look at the root config file
then                                # or user config file
conf=/etc/my.cnf
else
conf=$HOME/.my.cnf       # Using the users config file
fi

# The following code tries to get the variables safe_mysqld needs from the
# config file.This isn't perfect as this ignores groups, but it should
# work as the options doesn't conflict with anything else.

if test -f "$conf"            # Extract those fields we need from config file.
then
if grep "^datadir" $conf > /dev/null
then
datadir=`grep "^datadir" $conf | cut -f 2 -d= | tr -d ' '`
fi
if grep "^user" $conf > /dev/null
then
mysql_daemon_user=`grep "^user" $conf | cut -f 2 -d= | tr -d ' ' | head -1`
fi
if grep "^pid-file" $conf > /dev/null
then
pid_file=`grep "^pid-file" $conf | cut -f 2 -d= | tr -d ' '`
else
if test -d "$datadir"
then
pid_file=$datadir/`hostname`.pid
fi
```

```
fi
if grep "^basedir" $conf > /dev/null
then
basedir=`grep "^basedir" $conf | cut -f 2 -d= | tr -d ' '`
bindir=$basedir/bin
fi
if grep "^bindir" $conf > /dev/null
then
bindir=`grep "^bindir" $conf | cut -f 2 -d=| tr -d ' '`
fi
fi


# Safeguard（relative paths, core dumps.）
cd $basedir

case "$mode" in
'start')
# Start daemon

if test -x $bindir/safe_mysqld
then
# Give extra arguments to mysqld with the my.cnf file.This script may
# be overwritten at next upgrade.
$bindir/safe_mysqld --user=$mysql_daemon_user --pid-file=$pid_file
--datadir=$datadir &
else
echo "Can't execute $bindir/safe_mysqld"
fi
```

```
        ;;

'stop')
# Stop daemon.We use a signal here to avoid having to know the
# root password.
if test -f "$pid_file"
then
mysqld_pid=`cat $pid_file`
echo "Killing mysqld with pid $mysqld_pid"
kill $mysqld_pid
# mysqld should remove the pid_file when it exits.
else
echo "No mysqld pid file found.Looked for $pid_file."
fi
        ;;

    *)
# usage
echo "usage：$0 start|stop"
exit 1
        ;;
esac
```

以下举例显示了如何使用 cluadmin 来添加 MySQL 服务 。

```
cluadmin> service add

The user interface will prompt you for information about the service.
Not all information is required for all services.
```

```
Enter a question mark（?）at a prompt to obtain help.

Enter a colon（：）and a single-character command at a prompt to do
one of the following：

c - Cancel and return to the top-level cluadmin command
r - Restart to the initial prompt while keeping previous responses
p - Proceed with the next prompt

Currently defined services:

Service name：mysql
Preferred member [None]：test1
User script（e.g., /usr/foo/script or None）[None]：/etc/rc.d/init.d/mysql.server
Do you want to add a check script to the service (yes/no/?) [no]: yes

        Check Script Information

Check script (e.g., "/opt/cluster/usercheck/httpCheck 172.16.69.200 80" or None)
[None]: /home/oracle/mysqlCheck
Check interval (in seconds) [None]: 5
Check timeout (in seconds) [None]: 30
Max error count [None]: 8
Do you want to (m)odify, (d)elete or (s)how the check script, or are you (f)inished
adding check script: f
Do you want to add floating IP address to the service (yes/no/?) [no]: yes

        IP Address Information
```

```
IP address： 10.1.16.12
Net interface [None]: eth0
Netmask（e.g. 255.255.255.0 or None）[None]： 255.255.255.0
Broadcast（e.g. X.Y.Z.255 or None）[None]： 10.1.16.255
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you (f)inished adding IP addresses: f
Do you want to add a nbd mirror disk device to the service (yes/no/?) [no]: no
Do you want to add a disk device to the service（yes/no/?）：yes

Disk Device Information

Device special file（e.g., /dev/sda1）：/dev/sda1
Check service device (yes/no/?) [yes]: yes
Device is SCSI (yes/no/?) [yes]: yes
Device check timeout (in seconds) [120]:
Filesystem type（e.g., ext2, reiserfs, ext3 or None）：ext3
Mount point（e.g., /usr/mnt/service1 or None）[None]：/var/mysql
Mount options（e.g., rw, nosuid）：[Return] rw
Forced unmount support（yes/no/?）[no]：yes
Device owner（e.g., root）：root
Device group（e.g., root）：root
Device mode（e.g., 755）： 755
Do you want to (a)dd, (m)odify, (d)elete or (s)how devices, or are you (f)inished adding devices: f
Service start timeout (in seconds) [None]:
Service stop timeout (in seconds) [None]:
Reboot system if stop the service failed (yes/no/?) [yes]:
Disable service（yes/no/?）[no]：no
```

```
service name: mysql
disabled：no
preferred node：test1
user script：/etc/rc.d/init.d/mysql.server
check script: /home/oracle/mysqlCheck
   check interval: 5
   check timeout: 30
   max error count: 8
IP address 0：  10.1.16.12
   net interface 0: eth0
netmask 0：  255.255.255.0
broadcast 0：  10.1.16.255
device 0：/dev/sda1
   check device, device 0: yes
   SCSI device, device 0: yes
   check timeout, device 0: 120
mount point, device 0：/var/mysql
mount fstype, device 0：ext3
force unmount, device 0：yes
owner, device 0: root
group, device 0: root
mode, device 0: 755
start timeout: None
stop timeout: None
reboot system if stop service failed: yes
Add mysql service as shown? (yes/no/?) yes
   0) test1
   1) test2
   c) cancel
```

```
Choose member to start service on: 0
Added mysql.
cluadmin>
```

### 7.4.3 配置 DB2 服务

本节举例说明了如何设置一项集群服务，使其在 GreatTurbo HA 10 之上对 IBM DB2 Enterprise/Workgroup Edition 进行故障切换。 本举例假设 NIS 没有运行在集群系统上。

如果想在集群系统上安装软件和数据库，请遵循以下步骤：

- 在两个集群系统上，以 root 身份登录并添加将用来访问 DB2 服务中 /etc/hosts 文件的 IP 地址和主机名称。例如：

```
10.1.16.182        ibmdb2.class.cluster.com           ibmdb2
```

- 在某一共享磁盘上选择一个未使用的分区来托管 DB2 管理和例程数据，并在其上创建一个文件系统。例如：

```
# mke2fs /dev/sda3
```

- 在两个集群系统上为步骤 2 中所创建的文件系统创建一个加载点。例如：

```
# mkdir /db2home
```

➢ 在第一个集群系统（devel0）上，将步骤 2 中创建的文件系统加载到步骤 3 中创建的加载点上。例如：

```
devel0# mount -t ext2 /dev/sda3 /db2home
```

➢ 在第一个集群系统(devel0)上,加载 DB2 cdrom 并复制 distribution to /root 中所包含的设置响应文件。例如：

```
devel0% mount -t iso9660 /dev/cdrom /mnt/cdrom
devel0% cp /mnt/cdrom/IBM/DB2/db2server.rsp /root
```

➢ 修改设置响应文件 db2server.rsp，显示本地配置设置。确保 UID 和 GID 保存在两个集群系统上。例如：

```
-----------Instance Creation Settings------------
-------------------------------------------------
DB2.UID = 2001
DB2.GID = 2001
DB2.HOME_DIRECTORY = /db2home/db2inst1

-----------Fenced User Creation Settings----------
--------------------------------------------------
```

```
UDF.UID = 2000

UDF.GID = 2000

UDF.HOME_DIRECTORY = /db2home/db2fenc1


-----------Instance Profile Registry Settings------

--------------------------------------------------

DB2.DB2COMM = TCPIP


----------Administration Server Creation Settings---

---------------------------------------------------

ADMIN.UID = 2002

ADMIN.GID = 2002

ADMIN.HOME_DIRECTORY = /db2home/db2as


---------Administration Server Profile Registry Settings-

--------------------------------------------------------

ADMIN.DB2COMM = TCPIP


---------Global Profile Registry Settings-------------

-----------------------------------------------------

DB2SYSTEM = ibmdb2
```

- 启动安装。例如：

```
devel0# cd /mnt/cdrom/IBM/DB2

devel0# ./db2setup -d -r /root/db2server.rsp 1>/dev/null 2>/dev/null &
```

- 通过检查安装日志文件/tmp/db2setup.log 来查看安装过程中出现的错误。安装过程中的每一步都必须确保日志文件结尾处标记为 SUCCESS（成功）。

- 停止第一个集群系统上的 DB2 例程和管理服务器。例如：

```
devel0# su - db2inst1
devel0# db2stop
devel0# exit
devel0# su - db2as
devel0# db2admin stop
devel0# exit
```

- 卸载第一个集群系统上的 DB2 例程和管理数据分区。例如：

```
devel0# umount /db2home
```

- 在第二个集群系统（devel1）上加载 DB2 例程和管理数据分区。例如：

```
devel1# mount -t ext2 /dev/sda3 /db2home
```

- 在第二个集群系统上加载 DB2 cdrom 并将 db2server.rsp 文件远程复制到 /root。例如：

```
devel1# mount -t iso9660 /dev/cdrom /mnt/cdrom
devel1# rcp devel0：/root/db2server.rsp /root
```

- 在第二个集群系统（devel1）上启动安装。例如：

```
devel1# cd /mnt/cdrom/IBM/DB2
devel1# ./db2setup -d -r /root/db2server.rsp 1>/dev/null 2>/dev/null &
```

- 通过检查安装日志文件来查看安装过程中出现的错误。必须确保安装过程中的每一步都标记为 SUCCESS（成功），下列情况除外：

```
DB2 Instance Creation                              FAILURE
Update DBM configuration file for TCP/IP           CANCEL
Update parameter DB2COMM                           CANCEL
Auto start DB2 Instance                            CANCEL
DB2 Sample Database                                CANCEL
Start DB2 Instance
```

```
Administration Server Creation                    FAILURE
Update parameter DB2COMM                                CANCEL
Start Administration Serve                        CANCEL
```

➢ 通过调用下列命令先后测试两个集群系统上的数据库安装情况：

```
# mount -t ext2 /dev/sda3 /db2home
# su - db2inst1
# db2start
# db2 connect to sample
# db2 select tabname from syscat.tables
# db2 connect reset
# db2stop
# exit
# umount /db2home
```

➢ 在 DB2 管理和例程数据分区上创建 DB2 集群启动/停止脚本。例如：

```
# vi /db2home/ibmdb2
# chmod u+x /db2home/ibmdb2

#!/bin/sh
#
# IBM DB2 Database Cluster Start/Stop Script
```

```
#

DB2DIR=/usr/IBMdb2/V6.1

case $1 in
"start")
$DB2DIR/instance/db2istrt
    ;;
"stop")
$DB2DIR/instance/db2ishut
    ;;
esac
```

- 在停用数据库之前，修改两个集群系统上的/usr/IBMdb2/V6.1/instance/db2ishut 文件，以有效断开活动的应用程序。例如：

```
for DB2INST in ${DB2INSTLIST?}; do
echo "Stopping DB2 Instance "${DB2INST?}"...">> ${LOGFILE?}
find_homedir ${DB2INST?}
INSTHOME="${USERHOME?}"
su ${DB2INST?}-c " \
source ${INSTHOME?}/sqllib/db2cshrc     1> /dev/null 2> /dev/null; \
${INSTHOME?}/sqllib/db2profile 1> /dev/null 2> /dev/null; \
>>>>>>> db2 force application all; \
db2stop                                          "    1>>   ${LOGFILE?}2>>
```

```
${LOGFILE?}
if [ $?-ne 0 ]; then
ERRORFOUND=${TRUE?}
fi
done
```

- 编辑 inittab 文件并取消 DB2 行，使集群服务能处理启动和停止 DB2 服务。这通常是文件的最后一行。例如：

```
# db：234：once：/etc/rc.db2 > /dev/console 2>&1 # Autostart DB2 Services
```

使用 cluadmin 工具创建 DB2 服务。添加步骤 1 中的 IP 地址、步骤 2 中所创建的共享分区以及步骤 16 中所创建的启动/停止脚本。

如果想在第三个系统上安装 DB2 客户机，则调用以下命令：

```
display# mount -t iso9660 /dev/cdrom /mnt/cdrom
display# cd /mnt/cdrom/IBM/DB2
display# ./db2setup -d -r /root/db2client.rsp
```

如果想配置 DB2 客户机，则将服务的 IP 地址添加到客户机系统上的/etc/hosts 文件中。例如：

```
10.1.16.182     ibmdb2.lowell.mclinux.com     ibmdb2
```

然后，再将下列条目添加到客户机系统上的/etc/services 文件中：

```
db2cdb2inst1          50000/tcp
```

在客户机系统上调用以下命令：

```
# su - db2inst1
# db2 catalog tcpip node ibmdb2 remote ibmdb2 server db2cdb2inst1
# db2 catalog database sample as db2 at node ibmdb2
# db2 list node directory
# db2 list database directory
```

如果想测试 DB2 客户机系统的数据库，则调用以下命令：

```
# db2 connect to db2 user db2inst1 using ibmdb2
```

```
# db2 select tabname from syscat.tables

# db2 connect reset
```

### 7.4.4 配置 Apache 服务

本节举例说明了如何设置一项集群服务，使其对某台 Apache Web 服务器进行故障切换。尽管服务中使用的实际变量取决于您的具体配置，但本举例也可帮助您为自己的环境设置该服务。

如果想设置 Apache 服务，您必须将两个集群系统都配置为 Apache 服务器。该集群软件可确保每次只有一个集群系统运行 Apache 软件。

当您在集群系统上安装 Apache 软件时，不要配置集群系统，以使 Apache 在系统启动时能够自动启动。例如，如果您在运行级目录（比如/etc/rc.d/init.d/rc3.d）中使用 Apache，则 Apache 软件将同时在两个集群系统上启动，这可能导致数据遭到损坏。

当您添加某项 Apache 服务时，您必须为其指定一个 “ 浮动 ” IP 地址。集群基础设施会将该 IP 地址绑定到当前正在运行 Apache 服务的集群系统的网络接口上。该 IP 地址可确保运行 Apache 软件的集群系统对访问 Apache 服务器的 HTTP 客户机是透明的。

而当集群系统启动时，禁止将包含 Web 内容的文件系统自动加载到共享磁盘存储上。相反，当 Apache 服务在集群系统上被启动和停止时，集群软件必须相应地加载和卸载文件系统。这样可防止两个集群系统同时访问同一数据（可能导致数据损坏）。因此，不要在/etc/fstab 文件中包含文件系统。

设置 Apache 服务时需执行以下 4 个步骤：

1. 为服务设置共享文件系统。
2. 在两个集群系统上安装 Apache 软件。
3. 在两个集群系统上配置 Apache 软件。

4. 将服务添加到集群数据库中。

如果想为 Apache 服务设置共享文件系统，则以 root 身份登录并在其中一个集群系统上执行以下任务：

1. 在某个共享磁盘上使用交互式 fdisk 命令来创建一个将用于 Apache 文件根目录的分区。注：可在不同的磁盘分区上创建多个文件根目录。
2. 使用 mkfs 命令在您上一步骤所创建的分区上建立一个 ext2 文件系统。指定盘符和分区号码。例如：

```
# mkfs /dev/sde3
```

3. 在 Apache 文件根目录上加载将包含 Web 内容的文件系统。例如：

```
# mount /dev/sde3 /opt/apache-1.3.12/htdocs
```

4. 不要将该加载信息添加到/etc/fstab 文件中，因为只有集群软件才能在服务中加载和下载文件系统。
5. 将所有需要的文件复制到文件根目录下。
6. 如果您有必须存在不同目录下或是独立分区的 CGI 文件或其它文件，请根据需要重复以上步骤。

您必须在两个集群系统上安装 Apache 软件。注意，两个集群服务器上的基本 Apache 服务器配置必须相同，以便服务可正确地进行故障切换。以下举例说明了某台基本 Apache web 服务器的安装情况（不带第三方模块或未进行性能调整）。如果想安装带有模块的 Apache 或对其进行调整以获得更出色的性能，请参见 Apache 安装目录下的 Apache 文件或访问 Apache 网站：www.apache.org。

在两个集群系统上，按以下步骤安装 Apache 软件：

1. 获取 Apache 软件 tar 文件。改变到/var/tmp 目录，然后使用 ftp 命令访问 Apache ftp 镜像站点：ftp.digex.net。在站点内，更改到包含 tar 文件的远程目录，并使用 get 命令将文件复制到集群系统上，然后从 FTP 站点断开连接。例如：

```
# mount /dev/sde3 /opt/apache-1.3.12/htdocs
# cd /var/tmp
# ftp ftp.digex.net
ftp> cd /pub/packages/network/apache/
ftp> get apache_1.3.12.tar.gz
ftp> quit
#
```

2. 从 Apache tar 文件中析取所需的文件。例如：

```
# tar -zxvf apache_1.3.12.tar.gz
```

3. 更改到步骤 2 中所创建的 Apache 安装目录。例如：

```
# cd apache_1.3.12
```

4. 为安装 Apache 创建一个目录。例如：

```
# mkdir /opt/apache-1.3.12
```

5. 调用配置命令，指定您在步骤 4 中所创建的 Apache 安装目录。如果您想定制安装，则调用配置—帮助命令来显示可用配置选项，或读取 Apache 安装或 README 文件。例如：

```
# ./configure --prefix=/opt/apache-1.3.12
```

6. 建立并安装 Apache 服务器。例如：

```
# make
# make install
```

7. 先后为该组添加 group nobody 和 user nobody（除非它们已经存在）。接着将 Apache 安装目录的所有者改为 nobody。例如：

```
# groupadd nobody
# useradd -G nobody nobody
# chown -R nobody.nobody /opt/apache-1.3.12
```

如果想将集群系统配置为 Apache 服务器，应先定制 httpd.conf Apache 配置文件，并创建一个可启动和停止 Apache 服务的脚本。然后将文件复制到其它集群系统上。两个集群系统上的文件必须相同，以使 Apache 服务正确地进行故障切换。

在其中一个系统上执行以下任务：

1.编辑/opt/apache-1.3.12/conf/httpd.conf Apache 配置文件，并根据您的配置定制该文件。例如：

- 指定保持启用的最大请求数量：

```
MaxKeepAliveRequests n
```

- 采用适当值替换 n，该值至少为 100。如欲获得最佳性能，则将无限制请求指定为 0。
- 指定客户机的最大数量：

```
MaxClients n
```

- 采用适当值替换 n。在缺省情况下，您最多可以指定 256 台客户机。如果您需要更多客户机，则您必须重新编译支持更多客户机的 Apache。更多详情，请参见 Apache 文件。
- 指定 user nobody 和 group nobody。注意，进行设置时必须使之与 Apache 主目录和文件根目录下的权限相匹配。例如：

```
User nobody

Group nobody
```

- 指定将包含 HTML 文件的目录。当您将 Apache 服务添加到集群数据库时，必须指定此加载点。例如：

```
DocumentRoot "/opt/apache-1.3.12/htdocs"
```

- 指定将包含 CGI 程序的目录。例如：

```
ScriptAlias /cgi-bin/ "/opt/apache-1.3.12/cgi-bin/"
```

- 指定在上一步骤中所使用的路径，并为该目录设置缺省的访问权限。例如：

```
<Directory opt/apache-1.3.12/cgi-bin">

AllowOverride None
```

```
Options None
Order allow,deny
Allow from all
</Directory>
```

如果您想调整 Apache 或添加第三方模块功能，则可能需要作出其它修改。关于设置其它选项的信息，请参见 Apache 项目文件。

1. 标准 Apache 启动脚本可能不接受集群基础设施传递给它的参数，所以您必须创建一个服务启动和停止脚本，使其只把第一个参数传递给标准 Apache 启动脚本。执行此任务时，您必须创建/etc/opt/cluster/apwrap 脚本，使之包括以下行：

```
#!/bin/sh
/opt/apache-1.3.12/bin/apachectl $1
```

2. 注意，Apache 启动脚本的实际名称取决于 Linux 发行版。例如，文件名可能会是/etc/rc.d/init.d/httpd。
3. 修改步骤 2 中所创建的脚本的权限，以使其能够得到执行。例如：

```
chmod 755 /etc/opt/cluster/apwrap
```

4. 使用 ftp、rcp 或 scp 命令将 httpd.conf 和 apwrap 文件复制到其它集群系统上。

在将 Apache 服务添加到集群数据库之前，必须确保没有安装 Apache 目录。然后，在其中一个集群系统上添加该项服务。您必须指定一个 IP 地址，集群基础设施会将其绑定到集群系统（运行 Apache 服务）的网络接口上。

下面是对使用 cluadmin 工具添加 Apache 服务的举例。

```
cluadmin> service add

The user interface will prompt you for information about the service.
Not all information is required for all services.

Enter a question mark（?）at a prompt to obtain help.

Enter a colon（：）and a single-character command at a prompt to do
one of the following：

c - Cancel and return to the top-level cluadmin command
r - Restart to the initial prompt while keeping previous responses
p - Proceed with the next prompt

Currently defined services:

Service name：apache
Preferred member [None]：devel0
User script（e.g., /usr/foo/script or None）[None]：/etc/opt/cluster/apwrap
Do you want to add a check script to the service (yes/no/?) [no]: yes

        Check Script Information

Check script (e.g., "/opt/cluster/usercheck/httpCheck 172.16.69.200 80" or None)
[None]: /opt/cluster/usercheck/httpCheck 10.1.16.150 80
Check interval (in seconds) [None]: 30
Check timeout (in seconds) [None]: 20
Max error count [None]: 3
```

```
Do you want to (m)odify, (d)elete or (s)how the check script, or are you (f)inished adding check script: f
Do you want to add floating IP address to the service (yes/no/?) [no]: yes

        IP Address Information

IP address： 10.1.16.150
Net interface [None]: eth0
Netmask（e.g. 255.255.255.0 or None）[None]： 255.255.255.0
Broadcast（e.g. X.Y.Z.255 or None）[None]： 10.1.16.255
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you (f)inished adding IP addresses: f
Do you want to add a nbd mirror disk device to the service (yes/no/?) [no]: no
Do you want to add a disk device to the service（yes/no/?）：yes

Disk Device Information

Device special file（e.g., /dev/sda1）：/dev/sda3
Check service device (yes/no/?) [yes]: yes
Device is SCSI (yes/no/?) [yes]: yes
Device check timeout (in seconds) [120]:
Filesystem type（e.g., ext2, reiserfs, ext3 or None）：ext3
Mount point（e.g., /usr/mnt/service1 or None）[None]：/opt/apache-1.3.12/htdocs
Mount options（e.g., rw, nosuid）：[Return] rw,sync
Forced unmount support（yes/no/?）[no]：yes
Device owner（e.g., root）：nobody
Device group（e.g., root）：nobody
Device mode（e.g., 755）： 755
Do you want to (a)dd, (m)odify, (d)elete or (s)how devices, or are you (f)inished
```

```
adding devices: f
Service start timeout (in seconds) [None]:
Service stop timeout (in seconds) [None]:
Reboot system if stop the service failed (yes/no/?) [yes]:
Disable service（yes/no/?）[no]：no
service name: apache
disabled：no
preferred node：devel0
user script：/etc/opt/cluster/apwrap
check script: /opt/cluster/usercheck/httpCheck 10.1.16.150 80
   check interval: 30
   check timeout: 20
   max error count: 3
IP address 0：  10.1.16.150
   net interface 0: eth0
netmask 0：  255.255.255.0
broadcast 0：  10.1.16.255
device 0：/dev/sda3
   check device, device 0: yes
   SCSI device, device 0: yes
   check timeout, device 0: 120
mount point, device 0：/opt/apache-1.3.12/htdocs
mount fstype, device 0：ext3
mount options, device 0：rw,sync
force unmount, device 0：yes
owner, device 0: nobody
group, device 0: nobody
mode, device 0: 755
start timeout: None
```

```
stop timeout: None
reboot system if stop service failed: yes
Add apache service as shown? (yes/no/?) yes
  0) devel0
  1) devel1
  c) cancel

Choose member to start service on: 0
Added apache.
cluadmin>
```

### 7.4.5 配置 Domino 服务

IBM lotus domino 是一种高性能服务器，可支持工作流程和其它广泛的应用，包括 http、mail、dns 和 ftp 等。如果服务进行故障切换，则应用将通过新的集群系统访问共享数据。可通过网络访问的 domino 服务程序通常都指定有一个 IP 地址，该地址能够同这些服务一起进行故障切换，以保持客户机对它的透明访问。

本节举例说明了如何为 IBM lotus domino 设置集群服务。尽管服务脚本和其它脚本中使用的变量取决于具体的 domino 配置，但本举例仍可帮助您为自己的环境设置此项服务。

1. 准备工作

举例步骤如下：

- 对于 Domino 而言，通常要安装到两个目录中：programscl 目录和数据目录。它们在缺省情况下为/opt/lotus and /local/notesdata。我们通常使用下面的方法来安装集群：首先，为数据目录分配整个磁盘分区，这样即可使数据被集群中的不同成员所共享，也可让程序目录在不同的集群

成员中以完全相同的方式进行布局，您可以通过复制命令（如“ rcp –r ”）来实现这一步骤。这样可增强容错性，以防某个复制的 domino 执行程序为 damagedcl。另外，确保服务检查程序正确运行也非常必要。

在此举例中，/dev/sda11 被分配用于支持 domino 数据，其加载点为 /local/notesdata（Domino 的缺省数据路径）。

- 服务中包含一个供客户机使用的 IP 地址。例如，172.16.69.249。
- 非常重要！

如果希望domino服务检查程序能有效运行，则您必须复制3个domino配置文件：notes.ini、server.id 和 cert.id。在缺省情况下，它们位于数据目录（如/local/notesdata）中。我们建议您将这 3 个文件复制到{domino_program_directory}/notes/latest/linux，并始终与原文件保持一致，特别是在您对 donmino 配置做出修改之后。

比如，可将上述 3 个文件复制到/opt/lotus/notes/latest/linux 目录下。

- 如果希望客户机和服务检查程序访问 domino servicer，则您必须将 Domino 服务器与集群为其指定的同一虚拟 IP 地址链接起来。例如，/etc/hosts 文件包含以下项目。

```
* 172.16.69.248          domino.dev.cn.tlan      domino
```

- 如果想确保 domino 服务检查脚本正确运行，您必须在系统 lib 搜索路径上执行一个 domino lib--libnotes.so 。 libnotes.so 通常位于{domino_program_directory}/notes/latest/linux 目录下。您可以通过设置一个链接或使用 “ ldconfig ” 程序来完成这一步骤。比如，可以直接设置下面的链接。

```
ln -s /opt/lotus/notes/latest/linux/libnotes.so /usr/lib/libnotes.so
```

- 由于 Domino 服务器不带有集群软件所需的 Domino 服务启动/停止脚本，所以您需要自行创建一个。下面是一个与上例相对应的举例脚本，可供您参考。

```
#!/bin/sh
case "$1" in
'start')
su - notes -c /opt/lotus/bin/server > /dev/null 2> /dev/null &
exit 0
        ;;
'stop')
notes_pids=`ps -o pid -h -U notes`
kill -9 $notes_pids
exit 0
        ;;
   *)
echo "usage：$0 start|stop"
exit 1
        ;;
esac
```

上述所有项目都必须添加服务之前进行设置。

2. 将一项 domino 服务添加到集群。

   以下举例说明了如何使用 cluadmin 来添加某项 domino 服务 。

```
cluadmin> service add

The user interface will prompt you for information about the service.
Not all information is required for all services.

Enter a question mark（?）at a prompt to obtain help.

Enter a colon（：）and a single-character command at a prompt to do
one of the following：

c - Cancel and return to the top-level cluadmin command
r - Restart to the initial prompt while keeping previous responses
p - Proceed with the next prompt

Currently defined services:

Service name：domino
Preferred member [None]：
User script（e.g., /usr/foo/script or None）[None]：/etc/rc.d/init.d/domino
Do you want to add a check script to the service (yes/no/?) [no]: yes

        Check Script Information

Check script (e.g., "/opt/cluster/usercheck/httpCheck 172.16.69.200 80" or
None) [None]: /opt/ cluster/usercheck/dominoCheck domino
Check interval (in seconds) [None]: 30
Check timeout (in seconds) [None]: 20
Max error count [None]: 3
Do you want to (m)odify, (d)elete or (s)how the check script, or are you (f)inished
```

```
adding check script: f
Do you want to add floating IP address to the service (yes/no/?) [no]: yes

        IP Address Information

IP address： 172.16.69.248
Net interface [None]: eth0
Netmask（e.g. 255.255.255.0 or None）[None]： 255.255.252.0
Broadcast（e.g. X.Y.Z.255 or None）[None]： 172.16.71.255
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you
(f)inished adding IP addresses: f
Do you want to add a nbd mirror disk device to the service (yes/no/?) [no]: no
Do you want to add a disk device to the service（yes/no/?）：yes

Disk Device Information

Device special file（e.g., /dev/sda1）：/dev/sda3
Check service device (yes/no/?) [yes]: yes
Device is SCSI (yes/no/?) [yes]: yes
Device check timeout (in seconds) [120]:
Filesystem type（e.g., ext2, reiserfs, ext3 or None）：ext3
Mount point（e.g., /usr/mnt/service1 or None）[None]：/local/notesdata
Mount options（e.g., rw, nosuid）：
Forced unmount support（yes/no/?）[no]：yes
Device owner（e.g., root）：root
Device group（e.g., root）：root
Device mode（e.g., 755）： 755
Do you want to (a)dd, (m)odify, (d)elete or (s)how devices, or are you (f)inished
adding devices: f
```

```
Service start timeout (in seconds) [None]:
Service stop timeout (in seconds) [None]:
Reboot system if stop the service failed (yes/no/?) [yes]:
Disable service（yes/no/?）[no]：no
service name: domino
disabled：no
preferred node：none
user script：/etc/rc.d/init.d/domino
check script: /opt/cluster/usercheck/dominoCheck domino
   check interval: 30
   check timeout: 20
   max error count: 3
IP address 0：  172.16.69.248
   net interface 0: eth0
netmask 0：  255.255.252.0
broadcast 0：  172.16.71.255
device 0：/dev/sda3
   check device, device 0: yes
   SCSI device, device 0: yes
   check timeout, device 0: 120
mount point, device 0：/local/notesdata
mount fstype, device 0：ext3
force unmount, device 0：yes
owner, device 0: root
group, device 0: root
mode, device 0: 755
start timeout: None
stop timeout: None
reboot system if stop service failed: yes
```

```
Add domino service as shown? (yes/no/?) yes
   0) test1
   1) test2
   c) cancel

Choose member to start service on: 0
Added domino.
cluadmin>
```

### 7.4.6 配置 Informix 服务

数据库服务可为数据库应用（application：set）提供高可用性数据。之后，该应用即可为数据库客户机系统（比如 web 服务器等）提供网络访问能力。如果服务进行故障切换，则应用将通过新的集群系统访问共享数据库数据。可通过网络访问的数据库通常都指定有一个 IP 地址，该地址将同服务一起进行故障切换，以保持客户机的透明访问。

本节举例说明了如何为 Informix 数据库设置集群服务。尽管服务脚本中使用的变量取决于具体的 Informix 配置，但本举例也可帮助您为自己的环境设置某项服务。

1. 准备工作

   举例内容如下：

   - 添加 informix 服务之前，需要先在同一集群系统的两台服务器上创建一个带有 informix 名称的 informix 管理账户。
   - 在缺省的安装情况下，Informix 软件位于/opt/informix 中。
   - 安装后，在 informix 数据库系统可供使用之前，必须先初始化数据库服务器。通常，您应该编辑“ sqlhosts ”和“ onconfig ”配置文件（位于{informix-install-dir}/等目录下）。特别是，在没有“ onconfig ”

文件时，您可以从同一 placecl 中的模板文件“ onconfig.std ”中来创建它；如果您想更方便些，则可以使用“ onmonitor ”程序来创建。在缺省情况下，该程序位于{informix-install-directory}/bin 目录下。

在数据库初始化过程中，您至少需要设置那些与 Informix Root dbspace（用于容纳数据库）有关的属性。在举例中，Informix Root dbspace 位于整个磁盘分区/dev/sda9 目录下。

也就是说，磁盘分区由一个集群中的所有服务器共享；informix 程序可分别安装在每台服务器中，而在下次为集群配置服务时，可以将其安装在相同的目录下。

这种安排的好处在于可增强 informix 执行程序的容错性，且无需为集群中每台服务器的 Informix 管理员账号采用相同的组 id 和用户 id。

在另一种情形下，如果您将某一文件用作 Root dbspace（而非磁盘分区），则您应该让集群中每台服务器的 informix 管理员账号采用相同的组 id 和用户 id。

- informix 服务中包含一个供 informix 客户机使用的 IP 地址。例如 172.16.69.247。
- 为使 informix 服务检查脚本能正确运行，请检查{informix-install-directory}/等目录下的“ sqlhosts ”文件。例如，它可能像：

```
# dbservername          nettype             hostname
servicename         options
informixds1             onsoctcp            informixserver    sqlexec
```

- 确保主机名称（比如“ informixserver ”）与上述 ip 地址保持一致。所以您可能要将适当的项目添加到 etc/hosts 文件上，例如：

```
172.16.69.247          informixserver.dev.cn.tlan      informixserver
```

- 对环境来说，您应该设置 INFORMIXSERVER 和 INFORMIXDIR 环境变量并更新 PATH 环境变量。更多详情，请参见《Informix 指南》。例如，可按以下情况将它们添加在.bashrc 中：

```
export INFORMIXDIR=/opt/informix
export PATH=$PATH：$INFORMIXDIR/bin
export INFORMIXSERVER=informixds1
```

- 由于 informix 数据库软件不带有集群软件所需的 informix 服务启动/停止脚本，所以您需要自行创建一个。下面是一个与上例相对应的举例脚本，可供您参考。

```
#!/bin/sh
informix_install_dir=/opt/informix
if [ -n "$INFOMIXDIRi" -a -d $INFORMIXDIR ];
then
informix_install_dir=$INFORMIXDIR
fi
        case "$1" in
'start')
$informix_install_dir/bin/oninit
exit 0
;;
'stop')
```

```
$informix_install_dir/bin/onmode -ky

exit 0

;;

*)

echo "usage：$0 start|stop"

exit 1

;;

esac
```

上述所有项目都必须在添加服务之前进行设置。

2. 将一项 Informix 服务添加到集群。

   以下举例说明了如何使用 cluadmin 来添加某项 Informix 服务 。

```
Cluadmin> service add


The user interface will prompt you for information about the service.

Not all information is required for all services.


Enter a question mark（?）at a prompt to obtain help.


Enter a colon（：）and a single-character command at a prompt to do

one of the following：


c - Cancel and return to the top-level cluadmin command

r - Restart to the initial prompt while keeping previous responses

p - Proceed with the next prompt


Currently defined services:
```

```
Service name：infomix
Preferred member [None]：
User script（e.g., /usr/foo/script or None）[None]：/etc/rc.d/init.d/informix
Do you want to add a check script to the service (yes/no/?) [no]: yes

        Check Script Information

Check script (e.g., "/opt/cluster/usercheck/httpCheck 172.16.69.200 80" or None) [None]: /opt/cluster/usercheck/informixCheck informixds1 /opt/informix
Check interval (in seconds) [None]: 30
Check timeout (in seconds) [None]: 20
Max error count [None]: 3
Do you want to (m)odify, (d)elete or (s)how the check script, or are you (f)inished adding check script: f
Do you want to add floating IP address to the service (yes/no/?) [no]: yes

        IP Address Information

IP address： 172.16.69.247
Net interface [None]: eth0
Netmask（e.g. 255.255.255.0 or None）[None]： 255.255.252.0
Broadcast（e.g. X.Y.Z.255 or None）[None]： 172.16.71.255
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you (f)inished adding IP addresses: f
Do you want to add a nbd mirror disk device to the service (yes/no/?) [no]: no
Do you want to add a disk device to the service（yes/no/?）：no
Service start timeout (in seconds) [None]:
Service stop timeout (in seconds) [None]:
```

```
Reboot system if stop the service failed (yes/no/?) [yes]:
Disable service（yes/no/?）[no]：no
service name: infomix
disabled：no
preferred node：none
user script：/etc/rc.d/init.d/informix
check script: /opt/cluster/usercheck/informixCheck informixds1 /opt/informix
   check interval: 30
   check timeout: 20
   max error count: 3
IP address 0： 172.16.69.247
   net interface 0: eth0
netmask 0： 255.255.252.0
broadcast 0： 172.16.71.255
start timeout: None
stop timeout: None
reboot system if stop service failed: yes
Add infomix service as shown? (yes/no/?) yes
   0) test1
   1) test2
   c) cancel

Choose member to start service on: 0
Added infomix.
cluadmin>
```

### 7.4.7 配置 Sybase 服务

数据库服务可为数据库应用（application：set）提供高可用性数据。之后，该应用即可为数据库客户机系统（比如 web 服务器等）提供网络访问能力。如果服务进行故障切换，则应用将通过新的集群系统访问共享数据库数据。可通过网络访问的数据库通常都指定有一个 IP 地址，该地址将同服务一起进行故障切换，以保持客户机的透明访问。

本节举例说明了如何为 Sybase 数据库设置集群服务。尽管服务脚本中使用的变量取决于具体的 Sybase 配置,但本举例也可帮助您为自己的环境设置某项服务。

1. 准备工作

   举例内容如下：

   - 添加 sybase 服务之前，需要先在相同集群系统中的两台服务器上创建一个带有 sybase 名称的 sybase 管理账户。
   - 在缺省的安装情况下，Informix 软件位于/opt/sybase-11.9.2 下。
   - 安装后，在 sybase 数据库系统可供使用之前，您必须初始化数据库服务器。通常，您可以使用“ srvbuildi ”程序来完成这一步骤。在缺省情况下，该程序位于{Sybase-install-directory}/bin 目录下。

     在数据库初始化过程中，您至少需要设置那些与 Sybase 主数据库和 Sybsystemprocess 数据库有关的属性。在举例中，Sybase 主数据库被位于整个/dev/sda7 磁盘分区，而 Sybsystemprocess 数据库则位于另一个完整的磁盘分区/dev/sda10 上。

     也就是说，这两个磁盘分区由一个集群中所有的服务器所共享；sybase 程序可分别安装在每台服务器中，而在下一次为集群配置服务时，可以将其安装在相同的目录下。

这种安装的好处在于可增强 sybase 执行程序的容错性，且无需为集群中每台服务器的 Sybase 管理员账号采用相同的组 id 和用户 id。

在另一种情形下，如果您使用某一文件（而非磁盘分区）作为 Sybase 主数据库或 Sybase sybsystemprocess 数据库，则您应该让集群中每台服务器的 Sybase 管理员账号都采用相同的组 id 和用户 id。

- Sybase 服务中包含一个供 Sybase 客户机使用的 IP 地址。例如 172.16.69.250。
- 为使 sybase 服务检查脚本能正确运行，请检查位于{sybase-install-directory}目录下的 " interfaces " 文件（通过数据库初始化程序，比如" srvbuild "创建而成）。例如：sybase master tcp ether 172.16.69.250 4100 query tcp ether 172.16.69.250 4100 。确保数据库名称（比如 " sybase " ）与 sybase 服务检查脚本参数以及上述 ip 地址保持一致。
- 由于 Sybase 数据库软件带有 Sybase 服务启动/停止脚本，所以在完成数据库初始化后便可使用。
- 对环境来说，您最好将 SYBASE 环境变量设置为（syabse-install-directory），例如/opt/sybase-11.9.2。

上述所有项目都必须在添加服务之前进行设置。

2. 将一项 Sybase 服务添加到集群上。

以下举例说明了如何使用 cluadmin 来添加一项 Sybase 服务 。

```
Cluadmin> service add


The user interface will prompt you for information about the service.

Not all information is required for all services.
```

```
Enter a question mark（?）at a prompt to obtain help.

Enter a colon（：）and a single-character command at a prompt to do
one of the following：

c - Cancel and return to the top-level cluadmin command
r - Restart to the initial prompt while keeping previous responses
p - Proceed with the next prompt

Currently defined services:

Service name：sybase
Preferred member [None]：server1
User script （ e.g., /usr/foo/script or None ） [None] ：
/opt/sybase-11.9.2/install/rc.sybase
Do you want to add a check script to the service (yes/no/?) [no]: yes

        Check Script Information

Check script (e.g., "/opt/cluster/usercheck/httpCheck 172.16.69.200 80" or
None) [None]: /opt/cluster/usercheck/sybaseCheck sybase /opt/informix
Check interval (in seconds) [None]: 30
Check timeout (in seconds) [None]: 20
Max error count [None]: 3
Do you want to (m)odify, (d)elete or (s)how the check script, or are you (f)inished
adding check script: f
Do you want to add floating IP address to the service (yes/no/?) [no]: yes
```

```
            IP Address Information

IP address： 172.16.69.250
Net interface [None]: eth0
Netmask（e.g. 255.255.255.0 or None）[None]： 255.255.252.0
Broadcast（e.g. X.Y.Z.255 or None）[None]： 172.16.71.255
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you
(f)inished adding IP addresses: f
Do you want to add a nbd mirror disk device to the service (yes/no/?) [no]: no
Do you want to add a disk device to the service（yes/no/?）：no
Service start timeout (in seconds) [None]:
Service stop timeout (in seconds) [None]:
Reboot system if stop the service failed (yes/no/?) [yes]:
Disable service（yes/no/?）[no]：no
service name: sybase
disabled：no
preferred node：server1
user script：/opt/sybase-11.9.2/install/rc.sybase
check script: /opt/cluster/usercheck/sybaseCheck sybase
   check interval: 30
   check timeout: 20
   max error count: 3
IP address 0： 172.16.69.250
   net interface 0: eth0
netmask 0： 255.255.252.0
broadcast 0： 172.16.71.255
start timeout: None
stop timeout: None
reboot system if stop service failed: yes
```

```
Add sybase service as shown? (yes/no/?) yes
  0) server1
  1) server2
  c) cancel

Choose member to start service on: 0
Added sybase.
cluadmin>
```

### 7.4.8 配置 Websphere 服务

IBM Websphere 是一种高性能的 JSP/Servlet 服务器。它通常与其它 http 服务器（如 apache、netscape http 服务器等）一起在后台使用。如果服务进行故障切换，则应用将通过新的集群系统访问共享数据。可通过网络访问的 JSP/Servlet 服务通常都指定有一个 IP 地址，该地址将同服务一起进行故障切换，以保持客户机的透明访问。

本节举例说明了如何为 IBM WEbsphere 设置集群服务。尽管服务脚本中使用的变量取决于具体的 IBM Websphere 及其相关软件配置，但本举例也可帮助您为自己的环境设置某项服务。

1. 准备工作

   举例内容如下：

   - 由于 Websphere 没有明确地划分数据和程序，所以应将其安装在共享分区上。例如，/dev/sda11 被分配用于容纳 websphere 数据，而其加载点为/local/notesdata（Websphere 的缺省数据路径）。

- 为确保 Websphere 正确运行，您最好将 Webshere 与相应的 IBM 产品（比如 IBM HTTP 服务器和 IBM jdk1.1.8 等）一同使用。详细信息，请参见手册中的相关内容。
- Websphere 服务中包含一个供客户机使用的 IP 地址。例如，172.16.69.249。
- 非常重要！为使 Websphere 以相同的方式在两个集群节点中运行，您必须使用“ hostname ”命令为这些节点提供相同的主机名称，并将其与相同的虚拟 IP 地址相链接。例如，/etc/hosts 文件包含以下项目。

```
172.16.69.249        websphere.dev.cn.tlan    websphere
```

- 为确保 websphere 服务检查脚本正确运行，您必须确保“ snoop ” servlet 存在并运行于缺省主机之上。
- 由于 Websphere 服务器不带有集群软件所需的 Websphere 服务启动/停止脚本，所以您需要自行创建一个。下面是一个与上例相对应的举例脚本，可供您参考。

```
#!/bin/sh
case "$1" in
'start')
/etc/rc.d/init.d/ibmhttpd start
cd /opt/IBMWebAS/bin
startupServer.sh &
exit 0
        ;;
'stop')
killall -9 java 2> /dev/null
/etc/rc.d/init.d/ibmhttpd stop
```

```
exit 0
          ;;
    *)
echo "usage：$0 start|stop"
exit 1
          ;;
esac
```

上述所有项目都必须在添加服务之前进行设置。

2. 将一项 Websphere 服务添加到集群。

   以下举例说明了如何使用 cluadmin 来添加一项 Websphere 服务 。

```
Cluadmin> service add

The user interface will prompt you for information about the service.
Not all information is required for all services.

Enter a question mark（?）at a prompt to obtain help.

Enter a colon（：）and a single-character command at a prompt to do
one of the following：

c - Cancel and return to the top-level cluadmin command
r - Restart to the initial prompt while keeping previous responses
p - Proceed with the next prompt

Currently defined services:
```

```
Service name：websphere
Preferred member [None]：
User script（e.g., /usr/foo/script or None）[None]：/etc/rc.d/init.d/websphere
Do you want to add a check script to the service (yes/no/?) [no]: yes

        Check Script Information

Check script (e.g., "/opt/cluster/usercheck/httpCheck 172.16.69.200 80" or
None) [None]: /opt/cluster/usercheck/websphereCheck 172.16.69.249 80
Check interval (in seconds) [None]: 30
Check timeout (in seconds) [None]: 20
Max error count [None]: 3
Do you want to (m)odify, (d)elete or (s)how the check script, or are you (f)inished
adding check script: f
Do you want to add floating IP address to the service (yes/no/?) [no]: yes

        IP Address Information

IP address： 172.16.69.249
Net interface [None]: eth0
Netmask（e.g. 255.255.255.0 or None）[None]： 255.255.252.0
Broadcast（e.g. X.Y.Z.255 or None）[None]： 172.16.71.255
Do you want to (a)dd, (m)odify, (d)elete or (s)how an IP address, or are you
(f)inished adding IP addresses: f
Do you want to add a nbd mirror disk device to the service (yes/no/?) [no]: no
Do you want to add a disk device to the service（yes/no/?）：yes

Disk Device Information
```

```
Device special file（e.g., /dev/sda1）：/dev/sda6
Check service device (yes/no/?) [yes]: yes
Device is SCSI (yes/no/?) [yes]: yes
Device check timeout (in seconds) [120]:
Filesystem type（e.g., ext2, reiserfs, ext3 or None）：ext3
Mount point（e.g., /usr/mnt/service1 or None）[None]：/opt/IBMWebAS
Mount options（e.g., rw, nosuid）：rw
Forced unmount support（yes/no/?）[no]：yes
Device owner（e.g., root）：root
Device group（e.g., root）：root
Device mode（e.g., 755）： 755
Do you want to (a)dd, (m)odify, (d)elete or (s)how devices, or are you (f)inished
adding devices: f
Service start timeout (in seconds) [None]:
Service stop timeout (in seconds) [None]:
Reboot system if stop the service failed (yes/no/?) [yes]:
Disable service（yes/no/?）[no]：no
service name: websphere
disabled：no
preferred node：none
user script：/etc/rc.d/init.d/websphere
check script: /opt/cluster/usercheck/websphereCheck 172.16.69.249 80
   check interval: 30
   check timeout: 20
   max error count: 3
IP address 0： 172.16.69.249
   net interface 0: eth0
netmask 0： 255.255.252.0
broadcast 0： 172.16.71.255
```

```
device 0：/dev/sda6
   check device, device 0: yes
   SCSI device, device 0: yes
   check timeout, device 0: 120
mount point, device 0：/opt/IBMWebAS
mount fstype, device 0：ext3
mount options, device 0：rw
force unmount, device 0：yes
owner, device 0: root
group, device 0: root
mode, device 0: 755
start timeout: None
stop timeout: None
reboot system if stop service failed: yes
Add websphere service as shown? (yes/no/?) yes
   0) test1
   1) test2
   c) cancel

Choose member to start service on: 0
Added websphere.
cluadmin>
```

## 7.5 处理错误状态下的服务

当服务启动或停止失败时，服务将处在错误状态，处在错误状态下的服务，其资源的状态将无法确定（例如，它的部分资源已经释放，但某些服务资源仍然将在所有者节点上进行配置）。

GreatTurbo Cluster Server 10 中，以下的情形将导致服务的状态为 error：

- 启用服务失败
- 停止服务失败(多为服务检测失败时停止服务)
- 禁用服务失败
- 切换服务失败
- 停止 GreatTurbo Cluster Server 10 失败(即/etc/init.d/cluster stop 失败)

**注意：**

- **当服务在集群中任何一个节点上的状态为 error 时，将不能再启用服务。**
- **/etc/init.d/cluster stop 时，如果检查有服务的状态为 error，则会自动重启机器。**

由于处理错误状态下的服务可能有资源没有释放，所以处理错误状态中的服务必须十分小心。如果服务资源仍然在所有者节点上配置，那么在另一个集群节点上启动服务可能会造成严重的问题。例如，如果某一文件系统已经加载到它的所有者节点上，您又在另一个集群节点上启动该服务，则该文件系统将可能同时在两个系统上被加载，这样可能导致数据损坏。所以，您必须在确认所有的服务资源都已经完全释放掉了之后才能在另一个集群节点上重新启用服务。

GreatTurbo Cluster Server 10 对错误状态的服务的处理策略如下：

- 对启动服务和禁用服务失败的情况。由于这种操作时服务用户一般在现场，为了尽量避免机器重启动，所以启动服务失败时，资源的释放需要用户手动干预。

- 对停止服务和切换服务失败的情况。在配置服务时，用户可以指定在服务停止失败时是否重启机器，如果指定重启机器，则在服务停止失败时将自动重启机器，从而释放服务未释放的资源；如果没有指定重启机器，在服务停止失败时需要用户手动干预释放服务未释放的资源。

手动干预处理错误状态的服务的过程一般如下：

1) 定位错误的原因。

当服务的状态变为 error 时，首先要定位错误的原因，要从 GreatTurbo Cluster Server 10 的日志中分析。一般服务启动和停止的过程分为几步(例如，启动服务时会有 mount 分区，启动活动 IP，执行启动脚本等几步)，每一步出错都会打印日志，要从日志中分析具体是哪一步出错了。

2) 排除故障。

当找到错误的原因后，需要排除故障。例如，如果是启动/停止脚本写的有错误，则需要纠正错误并测试正确等。

3) 释放资源，恢复系统。

故障排除后，需要释放所有的没有释放的服务资源，并重新恢复系统。恢复的方法可分为两种情况：

- 机器允许重启。

这种情况处理比较简单，只需要手动 reboot 机器即可。

- 机器不允许重启。

很多情况下机器上面可能还有别的服务，因此不允许重启机器。在这种情况下，必须手动释放所有的资源(例如，手动 umont 磁盘设备等)，手动禁用服务，然后手动启用服务。

# 第八章 集群管理

在您设置某一集群并配置了服务之后，您可能需要管理集群，其中包括：

- 显示集群和服务状态
- 启动和停止集群软件
- 修改集群配置
- 备份和恢复集群数据库
- 修改集群事件日志
- 更新集群软件
- 重新加载集群数据库
- 修改集群名称
- 重新初始化集群
- 删除集群成员
- 诊断和纠正集群中的故障
- 图形管理和监控

## 8.1 使用文本界面进行集群管理

### 8.1.1 显示集群和服务的状态

监控集群和服务状态可帮您识别并解决集群环境中出现的问题。您可以通过使用下列工具来显示状态：

- cluadmin 工具
- clustat 命令

注意：状态是指您从集群中某一节点的角度看到的状态，如果您想全面了解集群状态，就应该在所有节点上都运行这一工具。

集群和服务状态包括以下信息：

- 集群成员系统状态
- 心跳通道状态
- 服务状态以及哪个集群系统在运行服务或拥有服务

下表说明了如何分析 cluadmin 工具、clustat 命令和集群 GUI 所显示的状态信息。

**表 1.**

| 成员状态 | 说明 |
|---|---|
| UP | 成员系统正在与另一成员系统通信和访问 quorum 分区。 |
| DOWN | 成员系统不能与另一成员系统通信。 |
| 心跳通道状态 | 说明 |
| OK | 心跳通道运行正常。 |
| Wrn | 不能获得通道状态。 |
| Err | 发生失败或错误。 |
| ONLINE | 心跳通道运行正常。 |
| OFFLINE | 虽然另一集群成员显示为 UP，但它在该通道上不能响应心跳请求。 |
| UNKNOWN | 不能获得该通道获得另一集群成员系统的状态，这可能是因为系统处于 DOWN 状态，或者集群守护进程没有运行。 |
| Service Status | 说明 |
| running | 服务资源已经配置，并可在包含该服务的集群系统上运行。running 状态 |

| | |
|---|---|
| | 是一种持久性状态。服务可从该状态进入 stopping 状态（例如，当首选成员重新加入集群时）、disabling 状态（当用户发出禁用该项服务的请求时）或 error 状态（当不能确定服务资源的状态时）。 |
| disabling | 服务处于被禁用过程中（例如，某一用户已经发出了禁用服务请求）。disabling 状态为过渡状态。在服务禁用成功或失败之前，服务将处于 disabling 状态。服务可从该状态进入 disabled 状态（如果禁用成功）、running 状态（如果禁用失败和服务被重启），或 error 状态（如果服务资源的状态不能确定）。 |
| disabled | 服务已经被禁用，而且不具有指定的所有者。disabled 状态为持久性状态。服务可从该状态进入 starting（正在启动）状态（当用户发出启动服务请求时）或 error 状态（当针对启动服务的请求失败并且服务资源状态不能确定时）。 |
| starting | 服务处于被启动过程中。starting 状态为过渡状态。服务在成功启动或启动失败之前将保持这种状态。服务可从该状态进入 running 状态（当服务启动成功时）、stopped 状态（如果服务停止失败），或 error 状态（当服务资源状态不能确定时）。 |
| stopping | 服务处于被停止过程中。stopping 状态为过渡状态。在服务被成功停止或停止失败之前，将处于 stopping 状态。服务可从该状态进入 stopped 状态（当服务被成功停止时）、running 状态（当服务停止失败并且能够启动时）或 error 状态（当服务资源的状态不能确定时）。 |
| stopped | 服务没有在任何集群系统上运行，没有一个指定的所有者，而且在集群系统上没有任何被配置的资源。stopped 状态是一种持久性状态。服务可从该状态进入 disabled（禁用）状态（当用户发出禁用服务请求时）或 starting 状态（当首选成员加入集群时）。 |
| error | 服务资源的状态不能确定。例如，某些与服务有关的资源可能仍需要在包含该服务的集群系统上进行配置。error 状态是一种持久性状态。为了保护数据的完整性，在您尝试启动或停止某一处于错误状态中的服务之前， |

| | |
|---|---|
| | 必须确保其服务资源已经不再集群系统上配置。 |
| not accept | 虽然某项服务资源的状态不能确定，但服务没有在该节点上运行。例如，与服务有关的某些资源可能仍然无效，比如服务绑定的网络接口失效。 |

如果想显示当前集群状态的快照，您可以调用其中一个集群系统上的 cluadmin 工具，并指定集群状态命令。 例如：

```
cluadmin> cluster status
Cluster Configuration (Turbolinux GreatTurbo HA 10):
Mon Apr 11 15:05:27 CST 2005

Member status:
    Member                  Id          System status
    ------------------- --------------- ---------------------
    test1                   0               Up
    test2                   1               Up

Channel status:
    Name                                    Type        Status
    ----------------------------------- ---------- ----------
    hb11 <--> hb21                       network     ONLINE
    hb12 <--> hb22                       network     ONLINE
    test1 <--> test2                     network     ONLINE
    /dev/ttyS0 <--> /dev/ttyS0           serial      ONLINE

Service status:
    Name                    On test1        On test2
    ------------------- --------------- ---------------------
    svc01                 stopped         running
```

```
cluadmin>
```

如果想监视集群并在每隔 5 秒显示状态快照，您可以指定 cluster monitor 命令。按 Return 或 Enter（回车）键停止显示。如果想修改时间间隔，您可以指定-interval time 命令（命令中的 time 指定的是状态快照的间隔秒数）。您可以在每次显示之后，通过指定-clear yes 命令选项来清除屏幕。缺省情况下将不清除屏幕。

如果只想显示集群服务的状态，您可以调用 cluadmin 工具并指定 service show state 命令。如果您知道希望显示的服务状态的名称，那么您可以指定 service show state service_name 命令。

您也可以使用 clustat 命令来显示集群和服务状态。如果想监视集群并显示具体时间间隔的状态,您可以选择-i time 命令来调用 clustat 程序(命令中的 time 指定的是状态快照的间隔秒数）。例如：

```
[root@test1 root]# /opt/cluster/bin/clustat –i 5
Cluster Configuration (Turbolinux GreatTurbo HA 10):
Mon Apr 11 15:05:27 CST 2005

Member status:
          Member                    Id          System status
          ------------------- -------------- --------------------
          test1                      0                Up
          test2                      1                Up

Channel status:
          Name                                    Type         Status
          ---------------------------------- ---------- ----------
          hb11 <--> hb21                       network      ONLINE
```

```
    hb12 <--> hb22                              network     ONLINE
    test1 <--> test2                            network     ONLINE
    /dev/ttyS0 <--> /dev/ttyS0                  serial      ONLINE

Service status:
    Name                        On test1        On test2
    ------------------- --------------- ---------------------
    svc01                       stopped         running
```

此外，您也可以使用 GUI 来显示集群和服务状态。更多详情，请参见**配置和使用图形用户界面**。

### 8.1.2 启动和停止集群软件

您可以通过调用 System V init 目录中的 cluster start 命令来启动一个集群系统上的集群软件。例如：

```
# /etc/rc.d/init.d/cluster start
```

您可以通过调用 System V init 目录中的 cluster stop 命令来停止一个集群系统上的集群软件。例如：

```
# /etc/rc.d/init.d/cluster stop
```

上述命令可能会导致集群系统的服务切换到其它集群系统上。

### 8.1.3 修改集群配置

您可能需要修改集群配置。例如，改正集群数据库中的心跳通道或 quorum 分区条目，复制的集群数据库位于/etc/opt/cluster/cluster.conf 文件中。

您必须使用 member_config 工具来修改集群配置，但不要修改 cluster.conf 文件。如果想修改集群配置，您必须按照**启动和停止集群软件**中的说明来停止其中一个集群系统上的集群软件。

然后调用 member_config 工具，并根据提示指定正确的信息。如果提示 wheather to run diskutil –I to quorum partion，则指定 no（否）。在运行程序后，重新启动集群软件。

### 8.1.4 备份和恢复集群数据库

建议您定期备份集群数据库。另外，您在对集群配置作出重大修改之前也应该备份数据库。

如果您想将集群数据库备份到/etc/opt/cluster/cluster.conf.bak 文件中，则请调用 cluadmin 工具，然后指定 cluster backup 命令。例如：

```
cluadmin> cluster backup
```

您也可以通过调用 cluadmin 工具并指定 cluster saveas filename 命令，将集群数据库保存到另外一个文件上。

如欲恢复集群数据库，请按以下步骤进行操作：

1. 通过调用 System V init 目录下 cluster stop 命令来停止其中一个系统上的集群软件。例如：

```
# /etc/rc.d/init.d/cluster stop
```

2. 上述命令可能会导致集群系统的服务故障切换到其它集群系统上。
3. 在另一个集群系统上，调用 cluadmin 工具并恢复集群数据库。如果想从 /etc/opt/cluster/cluster.conf.bak 文件恢复数据库，则指定 cluster restore 命令。如果想从另外一个文件恢复数据库，则指定 cluster restorefrom file_name 命令。

   集群将禁用所有正在运行的服务并予以删除，然后恢复数据库。

4. 通过调用 System V init 目录下的 cluster start 命令，来重启已停止的系统上的集群软件。例如：

```
# /etc/rc.d/init.d/cluster start
```

5. 先选择您希望在其上运行服务的集群系统，然后调用它的 cluadmin 工具并指定 service enable service_name 命令，启用每项集群服务。

### 8.1.5 修改集群事件日志

GreatTurbo Cluster Server 10 采用 Linux 的 syslog 系统来记录事件日志信息。

GreatTurbo Cluster Server 10 在安装时会自动编辑/etc/syslog.conf 文件，让集群将事件记录到 Linux 系统默认的日志文件/var/log/message 之外另一个日志文件中，GreatTurbo Cluster Server 10 默认的日志文件是/var/log/cluster。GreatTurbo Cluster Server 10 使用 syslogd 进程将有关的事件记录到/var/log/cluster 文件文件中。您可以使用日志文件来诊断集群中出现的问题。

GreatTurbo Cluster Server 10 的一个节点的日志文件只记录来自其所在节点的集群信息，所以您需要检查两个节点上的日志文件才能对集群的运行有全面的了解。

GreatTurbo Cluster Server 10 的事件日志可记录来自以下集群进程的信息：

- syncd – 数据通信进程
- svcmgr – 服务管理器进程
- powerd – 电源进程
- hb – 心跳进程
- svccheck – 服务检查进程
- clumon – 进程监视进程

事件的重要性决定于该事件的级别程度。重要事件应在其对集群可用性造成影响之前对其进行调查。GreatTurbo Cluster Server 10 的日志级别按照由高到低的顺序如下表所示：

| 值 | 严重性 | 描述 |
|---|---|---|
| 0 | emerg | 集群系统不可用 |
| 1 | alert | 必须立即针对该问题采取措施 |
| 2 | crit | 严重错误 |
| 3 | err | 错误 |
| 4 | warning | 发生了需要注意的重要事件 |
| 5 | notice | 发生了不影响系统运行的事件 |
| 6 | info | 正常操作的提示信息 |
| 7 | debug | 调试性信息 |

GreatTurbo Cluster Server 10 默认的日志级别为 info，系统中级别为 info 或级别比 info 高的日志都会记录在日志文件中。

您可以使用 cluadmin 或 guiadmin 工具修改 GreatTurbo Cluster Server 10 进程所记录日志的级别。在修改时只要输入日志级别所对应的整数值就可以了。

下面的例子介绍如何利用 cluadmin 工具将 syncd 进程记录的日志级别改为 debug：

```
[root@test1 root]# cluadmin
Thu Jul  7 17:32:32 CST 2005

You can obtain help by entering help and one of the following commands:

cluster     service     clear
help        apropos     nbd
exit
cluadmin> cluster loglevel syncd 7
cluadmin>
```

如果长时间运行后，/var/log/cluster 文件变得很大，可以使用命令 /opt/cluster/bin/clusterclear 将日志文件进行备份，并产生一个新的空的日志文件。

### 8.1.6 更新集群软件

您可以更新集群软件，但同时必须保存现有的集群数据库。更新某一系统上的集群软件可能要花 10 到 20 分钟，这取决于您是否需要重建内核程序。

如果您想在更新集群软件的同时最大限度地减少服务停机时间，则请遵循以下步骤进行操作：

1. 在您想更新的集群系统上，运行 cluadmin 工具并备份当前的集群数据库。例如：

```
cluadmin> cluster backup
```

2. 重新定位在第一个集群系统（您希望更新的系统）上运行的服务。更多详情，请参见**重新定位服务**。
3. 通过调用 System V init 目录下的 cluster stop 命令来停止第一个集群系统（您希望更新的系统）上的集群软件。例如：

```
# /etc/rc.d/init.d/cluster stop
```

4. 按照**安装和初始化集群软件**步骤中的说明，在您希望更新的第一个集群系统上安装最新的集群软件。当 member_config 工具提示您是否使用现有的集群数据库时，请选择 yes。
5. 通过调用 System V init 目录下的 cluster stop 命令来停止您希望更新的第二个集群系统上的集群软件。此时，将没有任何服务可用。
6. 通过调用 System V init 目录下的 cluster start 命令来启动第一个被更新的集群系统上的集群软件。此时，所有服务将变得可用。
7. 按照**安装和初始化集群软件**步骤中的说明，在您希望更新的第二个集群系统上安装最新的集群软件。当 member_config 工具提示您是否使用现有的集群数据库时，请指定 yes。
8. 通过调用 System V init 目录下的 cluster start 命令来启动第二个被更新集群系统上的集群软件。

### 8.1.7 重新加载集群数据库

调用 cluadmin 工具并使用集群 cluster reload 命令让集群重新读取集群数据库。例如：

```
cluadmin> cluster reload
```

### 8.1.8 修改集群名称

调用 cluadmin 工具并使用 cluster name cluster_name 命令为集群指定一个名称。集群名称在显示 clustat 命令和 GUI 中使用。例如：

```
cluadmin> cluster name cluster_1

cluster_1
```

### 8.1.9 重新初始化集群

在极少数情况下，您可能需要重新初始化集群系统、服务和数据库。但在重新初始化集群之前，您必须备份集群数据库。更多详情，请参见**备份和恢复集群数据库**。

如欲重新初始化集群，请遵循以下步骤进行操作：

1. 禁用所有正在运行的集群服务。
2. 通过调用 System V init 目录下的 cluster stop 命令来停止两个集群系统上的集群 daemons。例如：

   ```
   # /etc/rc.d/init.d/cluster stop
   ```

3. 在两个集群系统上安装集群软件。更多详情，请参见**安装和初始化集群软件**的步骤。
4. 在一个集群系统上运行 member_config 工具。当系统提示您是否使用现有的集群数据库，请指定 no。当系统提示您是否运行 diskutil –I 对 quorum 分区执行初始化时，请指定 yes。这样将从 quorum 分区中删除所有的状态信息和集群数据库。
5. 当 member_config 完成后，使用 scp 或 ftp 将配置文件复制到另一个集群系统上。

```
# scp /etc/opt/cluster/cluster*.conf server2：/etc/opt/cluster/
```

6. 在另一个集群系统上运行 member_config 工具。当系统提示您是否使用现有的集群数据库时，请选择 yes。当系统提示您是否运行 diskutil –I 对 quorum 分区执行初始化时，请指定 no。
7. 通过调用两个集群系统上 System V init 目录下的 cluster star 命令来启动集群 daemons。例如：

```
# /etc/rc.d/init.d/cluster start
```

### 8.1.10 删除集群中的成员

在某些情况下，您可能希望暂时删除集群中的某个成员系统。比如，当某个成员系统遇到硬件故障时，您可能需要重启系统，但为了在系统上执行维护任务，您必须防止该成员系统重新加入集群。

如果您正在运行某个 Red Hat 发行版，那么您可以使用 chkconfig 工具来做到在启动某个集群系统时不会使之重新加入到集群。例如：

```
# chkconfig --del cluster
```

您可以使用以下命令把系统重新加入集群：

```
# chkconfig --add cluster
```

如果您正在运行某个 Debian 发行版，则可以使用 update-rc.d 工具来启动某个集群系统，而不会使之重新加入集群。例如：

```
# update-rc.d -f cluster remove
```

您可以使用以下命令把系统重新加入集群：

```
# update-rc.d cluster defaults
```

这样，您就可以重启系统或是运行 System V init 目录下的 cluster start 命令了。例如：

```
# /etc/rc.d/init.d/cluster start
```

### 8.1.11 修改集群 watchdog 的超时时间

修改 watchdog 的超时时间,如果是软件 watchdog 时可以控制系统的负载的作用，如果是硬件 watchdog 时，可以控制系统死机后多长时间系统被 reboot。GreatTurbo Cluster Server 10 默认的 watchdog 超时时间是 600 秒。

在 cluadmin 管理器中，输入 “ cluster watchdog wdtimeout 超时时间数值 ”，就可以在线修改 watchdog 的超时时间。

```
cluadmin> cluster watchdog wdtimeout 600
cluadmin> cluster watchdog wdtimeout 1
Error: wdtimeout value must be between 5 and 3600
```

超时时间的数值必须在 5 秒和 3600 秒之间，如果不在这个范围，系统会提示出错，设置就不会生效。

### 8.1.12 修改集群的心跳属性

GreatTurbo Cluster Server 的心跳模块有三个属性：

- Hb interval：指心跳同步的时间间隔。单位是秒，取指范围 3 到 60 秒。默认值是 5。修改可以在线生效。

- Hb tko_count :指确认心跳失去连接的次数，也就是说，当心跳在 tko_count 次的时间间隔内，一直都不能联系的话，系统就会认为心跳通道故障。单位是次，取值范围 1 到 60 次。默认值是 3。修改可以在线生效。

- Hb port： 指 HA 模块 daemon 建立心跳连接时的端口号。取值范围 1024 到 65535。默认值是 1120。修改后，cluster 必须重新停止、启动后才能生效。

在 cluadmin 管理器中，可以按照如下办法修改上述参数：

```
cluadmin> cluster heartbeat interval 3
cluadmin> cluster heartbeat interval
3
cluadmin> cluster heartbeat interval 100
Error: heartbeat interval value must be between 2 and 60 seconds

cluadmin> cluster heartbeat tko_count 5
cluadmin> cluster heartbeat tko_count
5
cluadmin> cluster heartbeat tko_count 100
Error: heartbeat tko_count value must be between 1 and 60 times

cluadmin> cluster heartbeat port 1120
cluadmin> cluster heartbeat port
1120
cluadmin> cluster heartbeat port 100
Error: heartbeat port value must be between 1024 and 65535
```

### 8.1.13 修改集群告警邮件属性

在 GreatTurbo Cluster Server 10 所在系统发生故障时，GreatTurbo Cluster Server 会自动将错误信息发送给管理员的邮箱。使用该功能的前提是 GreatTurbo Cluster Server 所在的节点能够和同一内部网络的 smtp 服务进行正常连接，并且无需身份认证。

集群告警邮件的属性有四个：

- mail from：指发件人的邮箱地址，由于是 GreatTurbo Cluster Server 自动发送的地址，所以此处的发件人地址，可以输入该系统的名字，以和其他系统进行区别，必须按照邮箱地址的格式输入。例如：ha_oracle_server@bj_ha.com，后缀可以自行指定。
- mail to: 指收件人的邮箱地址，一般输入系统管理员的真实的邮件地址即可，这样当系统出现故障时，告警信息会自动发送给管理员的邮箱。
- smtpserver：指发送邮件的 smtp 服务器，可以是 ip 地址，也可以是 smtp 服务器的域名名称。
- mail level：指日志信息级别比 mail level 高的信息会发送邮件，值越小，级别越高。默认值是 4（LOG_ERR），表示当发送错误以上级别的信息时，发送邮件。**注意不要将该值改动过大，否则会导致不重要的信息也会发送邮件。**

仅当上述四个属性都进行设置后，邮件告警功能才会生效。系统默认没有设置邮件告警功能。可以按照如下方法：

```
cluadmin> cluster mail from ha_oracle_server@bj_ha.com
cluadmin> cluster mail to youraddr@yourcompany.com
cluadmin> cluster mail smtpserver mailservername
```

```
cluadmin> cluster mail level 4

cluadmin>
```

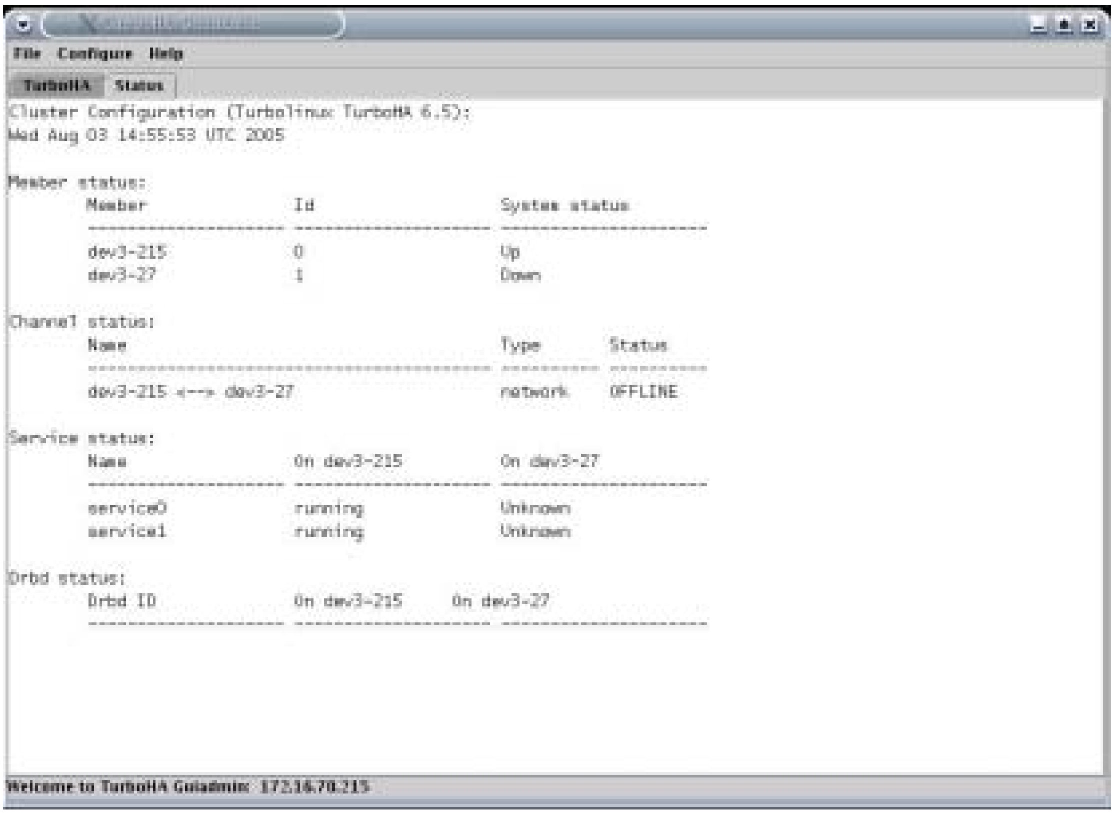
# 8.2 使用图形界面进行集群管理

通过 GreatTurbo Cluster Server 10 提供的 Guiadmin 图形工具，可以方便的进行集群的管理。目前，Guiadmin 不支持对 LB 服务的管理。

## 8.2.1 显示集群和服务的状态

guiadmin 中的状态模块用于给出集群中所有元素的快照视图。监控集群和服务状态可帮您识别并解决集群环境中出现的问题。您可以通过如下操作来显示状态：

点击标签栏中的“ Status ”标签，这样就可以显示集群和服务的状态了。状态信息从服务器端获取，每隔 3 秒钟刷新一次。




注意：状态是指您建立连接的那个节点的角度看到的状态。

集群和服务状态包括以下信息：

- 集群成员系统状态
- 心跳通道状态
- 服务状态以及哪个节点在运行服务或拥有服务

下表描述了 Status（状态）窗口的各项特性。

| 字段 | 说明 |
|---|---|
| Name | 显示集群的名称。 |
| Date | 集群状态快照的时间戳。 |
| Member Status | 显示同集群有关的所有成员的当前状态。 |
| Channel Status | 显示同集群有关的所有心跳通道的当前状态。 |
| Service Status | 显示了同集群有关的所有服务的当前状态。 |

### 8.2.2 如何在集群系统上运行 guiadmin

首先在集群系统上启动 X Window 系统，然后直接运行 guiadmin 客户端程序，并和本地或者对方节点 ip 连接即可以使用。由于 X Windows 系统需要占用大量的 CPU 和内存，所以在集群正常运行过程中不推荐您使用这种方法。

### 8.2.3 如何从远程系统上运行 guiadmin

首先按照快速安装文档正确的安装好 GreatTurbo Cluster Server 10 软件，把服务器端软件装在 GreatTurbo Cluster Server 10 两个节点上，把客户端软件装在远

程机器上；并且在 GreatTurbo Cluster Server 10 两节点上把 guiadmin 的服务器端进程启动。

然后在远程机器上运行 guiadmin 客户端程序，按照安装文档的说明连接至集群的任一节点。

这样，用户就可以远程对 HA 系统进行配置。

### 8.2.4 配置、修改心跳（heartbeat）参数

客户端软件成功连接到 Server 以后，会在配制模块显示出系统配置信息（图例见上一章）。

用户可以通过修改右侧面板中“ Heartbeat Parameters ”中的数据来修改 GreatTurbo Cluster Server 10 心跳参数。 参数的具体含义如下：

| 字段 | 说明 |
|---|---|
| Inteval | 心跳的时间间隔，以秒为单位。默认为 3 秒。此项值的范围是[2-60]秒。 |
| Port | 心跳链路的端口号。此项值的范围是[1024-65535]。 |
| TKO Count | 设定心跳连接允许的最大“ 失败 ”次数。此项值的范围是[1-60]。 |
| Third Part IP | 第三方 ip 地址，用来增强 HA 系统的稳定性。此处设置为只读。第三方 ip 地址使用 member_config 工具配置。 |

修改完毕以后，需要点击“ Apply ”键才能提交给服务器，使改动生效。

您也可以在客户端对其他模块进行修改，然后再点击“ Apply ”键，使所有改动一起提交给服务器。

### 8.2.5 配置、修改集群进程日志级别

用户通过修改配制模块右侧面板中 “ Log Level ” 的数据来修改 GreatTurbo Cluster Server 10 系统进程的日志级别，大于指定级别的日志信息将不写入日志文件中。 参数的具体含义如下：

| 字段 | 说明 |
|---|---|
| Svcmgr Log Level | 设置 Svcmgr 进程的日志级别。 |
| Powered Log Level | 设置 Power 进程的日志级别。 |
| Clumon Log Level | 设置 Clumon 进程的日志级别。 |
| guiadmin Log Level | 设置 guiadmin 进程的日志级别。 |
| Syncd Log Level | 设置 Syncd 进程的日志级别。 |
| Svccheck Log Level | 设置 Svccheck 进程的日志级别。 |
| Heartbeat Log Level | 设置心跳进程的日志级别。 |

修改完毕以后，需要点击 “ Apply ” 键才能提交给服务器，使改动生效。

您也可以在客户端对其他模块进行修改，然后再点击 “ Apply ” 键，使所有改动一起提交给服务器。

### 8.2.6 配置、修改邮件提示参数

用户通过修改配置模块右侧面板中 “ Mail Notify ” 中的数据来修改 GreatTurbo Cluster Server 10 系统中通知邮件的设置。当系统记录的日志级别大于用户指定的发送邮件级别时，将给用户指定的邮箱发送邮件。每个字段的具体含义如下：

| 字段 | 说明 |
|---|---|
| Mail from | 设置通知邮件的发送地址。 |

| 字段 | 说明 |
|---|---|
| Mail to | 设定通知邮件的接收地址。 |
| SMTP Server | 设置发送通知邮件的 smtp 服务器。 |
| Mail Level | 设定触发通知邮件发送的日志级别。 |

修改完毕以后，需要点击“ Apply ”键才能提交给服务器，使改动生效。

您也可以在服务器端对其他模块进行修改，然后再点击“ Apply ”键，使所有改动一起提交给服务器。

### 8.2.7 配置、修改 watchdog 信息

用户通过修改配置模块右边面板中“ Watchdog Parameters ”中的数据来修改 GreatTurbo Cluster Server 10 系统中 Watchdog 的设置。每个字段的具体含义如下：

| 字段 | 说明 |
|---|---|
| Watchdog Timeout | 设置 Watchdog 的最长等待时间。此项值的范围是[5-3600]秒。 |
| Member0 watchdo Device Name | 第一个节点的 watchdog 的设备名。 此项设置为只读。 |
| Member0 watchdo Device Name | 第二个节点的 watchdog 的设备名。此项设置为只读。 |

修改完毕以后，需要点击“ Apply ”键才能提交给服务器，使改动生效。

您也可以在 client 端对其他模块进行修改，然后再点击“ Apply ”键，使所有改动一起提交给服务器。

### 8.2.8 显示节点的配置信息

点击节点树中的 “ Members ”项，两个节点的配置信息会出现在右侧的主页面上。第一个节点信息显示在上半部，第二个节点信息显示在下半部。节点信息全部设置为只读，不能够修改。所有信息在 member_config 里面配置。在这个页面上用户可以查看以下信息：

1：成员信息，包括成员 id，成员的 hostname；

2：电子开关信息。如果没有配置电子开关，则此项为空。

3：心跳通道的信息。包括心跳的 id，名称（name），类型（type）和设备名（device）。如果用户配置的心跳类型为网络（net），那么 device 项目为空；如果用户配置的心跳类型为串口（serials），那么 device 项为串口的设备名。

下表中介绍了 Members 窗口中的各项特性。

**表 8. 配置成员字段**

| 字段 | 说明 |
| --- | --- |
| Member Info | |
| Member ID | 各成员唯一的节点号。 |
| Member Name | 成员的主机名。 |
| Power Switch Info | |
| Serial Port | 电子开关的端口号 |
| Switch type | 电子开关类型 |
| Heartbeat Channel Data | |
| ID | 心跳通道的 ID 号码 |
| Name | 心跳通道的名字。 |

| 字段 | 说明 |
|---|---|
| Type | 心跳通道的类型：网络或串行。 |
| Device | 串行连接的设备文件名。 |

在节点树上点击 Member0 或者 Member1 项可以查看每个节点的信息。这里的信息也全部为只读，用户不能够修改数据。

# 第九章　集群的维护

如果您还有什么疑问或者是您认为 GreatTurbo Cluster Server 10 在运行中出现问题，可以首先对照 FAQ 进行自行处理。如果发现仍然不能处理相应问题。可以与拓林思软件有限公司进行联系。

## 9.1 GreatTurbo Cluster Server 10 的日志信息

GreatTurbo Cluster Server 10 的日志信息存放在两个节点的/var/log/cluster 文件中，如果您的系统出现问题，您可以首先阅读两个节点的/var/log/cluster 日志文件，通过日志可以初步定位一些基本的故障和原因。

**注意：**

1) **在 GreatTurbo Cluster Server 10 运行期间，请不要手动修改和删除两个节点上的/var/log/clluster 文件，以免 log 信息丢失。**
2) **如果您发现/var/log/cluster 日志文件较大，可以在两个节点上分别直接运行“ /opt/cluster/bin/clusterclear ” 命令来备份并清空日志文件。**
3) **如果您发现/var/log/cluster 日志文件没有正常记录日志信息，可以执行“ /etc/init.d/syslog restart ”重新激活/var/log/cluster 文件的日志功能。**

## 9.2 Log 收集工具的使用方法

当 GreatTurbo Cluster Server 发生故障时，可以使用系统附带的 log 收集工具 clulogc 收集 log 和系统的信息，将收集到的信息发送给拓林思公司，以便更有效地定位故障的原因。

Log 收集工具的存放位置：/opt/cluster/bin/clulogc

Log 收集工具的使用方法：/opt/cluster/bin/clulogc　[OPTIONS]

| | | |
|---|---|---|
| 命令选项 | -a | 收集所有的信息 |
| | -t | 收集 GreatTurbo HA 相关的信息 |
| | -s | 收集 syslog |
| | -c | 收集 core 文件 |
| | -o | 收集系统的信息 |
| | -h | 显示帮助信息 |
| 参数选项 | -n clulog_rotate_num | 指定收集 GreatTurbo HA log 的文件个数 |
| | -p syslog_path | 指定 syslog 的路径 |
| | -r syslog_rotate_num | 指定收集 syslog 的文件个数 |

至少要指定一个命令行选项，参数选项可以指定也可以不指定。

当不指定任何参数选项时，表示收集所有的 GreatTurbo HA 的 log 和 syslog，并且 syslog 的路径为/var/log/message。

如果 syslog 的存放位置不在/var/log/message，需要利用-p 选项指定 syslog 的完整的路径。

假设要收集所有的 log 信息，并且只收集当前的 GreatTurbo HA 的 log 和 syslog，命令行应为：

/opt/cluster/bin/clulogc -a -n 1 -r 1

假设只收集 GreatTurbo HA 相关的所有信息，命令行应为：

/opt/cluster/bin/clulogc -t

收集到的 log 信息的存放位置：/opt/cluster/log/

收集到的 log 信息的文件名格式：机器名-年-月-日.tar.gz

例如机器名为 test1，收集日期为 2005 年 1 月 1 日，收集到的 log 信息的文件名为：test1-2005-01-01.tar.gz

**注意：**

1) **log 收集工具需要在两个节点上分别执行。**
2) **为了收集比较全面的信息，请用/opt/cluster/bin/clulogc –a 来收集信息；如果用-a 选项收集到的 log 文件过大，请分别用-t, -s, -c, -o 选项来收集信息。**

## 9.3 FAQ

1) **Q:** 如何查看 GreatTurbo Cluster Server 10 安装的是哪一个发行版本？

**A:** 可以使用 rpm -qa |grep greatturbocluster HA 命令来查看具体的版本号。

2) **Q:** 如果配置了第三方 IP 地址，但是第三方 IP 地址由于某种原因发生了一段时间(比如说 1 天)的故障,在这端时间内是否会影响 GreatTurbo Cluster Server 10 的正常使用？

**A:** 在这段时间内，如果没有节点发生故障，则不会影响 GreatTurbo Cluster Server 10 的正常使用，否则由于一个节点的故障（例如节点掉电或死机等），可能会导致另一个节点的反复 reboot 现象，从而会导致用户的业务受到影响。所以，如果不能保证第三方参考 IP 永久持续有效，则可以不配置第三方 IP。

3) **Q:** 如果不配置第三方 IP 地址，会对 GreatTurbo Cluster Server 10 系统有影响吗？

**A:** 如果两个节点之间配置了三条以上的心跳通道，那么这三条心跳通道全部故障的几率是非常小的。在这个前提下，如果用户实在是不具备永久持续有效的第三方参考 IP，则可以不配置第三方 IP 地址，在使用 member_config 工具的配置过程中将第三方 IP 配置选项选择“ no ”。如果用户具备永久持续有效的第三方参考 IP 时，则最好还是配上，在 member_config 的配置过程中将第三方 IP 配置选项选择“ yes ”，并输入地址值。

4) **Q:** 什么是永久持续有效的第三方参考 IP？

**A**：a）该永久持续有效，不会 down 掉，也不会短时故障。

b）从 HA 的节点能够 ping 通该 IP。

5) **Q:** GreatTurbo Cluster Server 10 为什么要配置多个心跳通道？

**A:** GreatTurbo Cluster Server 10 配置三条以上的通道是出于最大可能提高系统可用性的考虑。通过三条以上的通道来保证硬件上的更高冗余性。GreatTurbo Cluster Server 10 使用通道来同步心跳信息和发送命令。如果第一条通道正常，则 GreatTurbo HA.10 会选择第一条通道进行通讯，如果第一条通道故障，那么 GreatTurbo Cluster Server 10 会选择下一条正确的通道进行通讯，以此类推。也就是说 TuboHA10 按照 member_config 工具配置通道时的顺序来使用通道。

6) **Q:** GreatTurbo Cluster Server 10 除了将直连网线和串口配置成通道之外，为什么还需要将服务使用的外网网卡也配置成通道？

**A**：这是为了进一步保证服务的高可用性，以避免当所有的直连网线和串口都发生故障时会发生的裂脑现象，以保证应用数据的一致性。

需要强调的是，将服务使用的外网网卡配置成通道并不会影响服务的网络带宽，因为 GreatTurbo HA 会自动优先使用直连网线通道。只有所有的直连网线都发生故障时，GreatTurbo HA 才会使用外网网卡。

7) **Q:** 如何升级 GreatTurbo Cluster Server 10？

**A:** 如果由于某种原因，用户需要升级 GreatTurbo Cluster Server 10 到更高发行版本时，只需先按照手册卸载旧版本的 GreatTurbo Cluster Server 10，然后按照手册重新安装新版本的 GreatTurbo Cluster Server 10 即可。基本顺序是：1.卸载旧版本；2.安装新版本；3.如果原先配置的应用不需要更改或者也不需要增加新的应用，则无需重新配置，直接可以启动 GreatTurbo Cluster Server 10 进行使用。

8) **Q:** 如果用户的应用的 CPU 负载较重时，需要调整 GreatTurbo Cluster Server 10 的参数吗？

**A:** 如果用户的应用的 CPU 负载较重，且持续在 n*5 秒钟以内都维持在 100%以上时，需要使用 cluadmin 工具来进行调整 GreatTurbo Cluster Server 10 同步心跳的参数。进入 cluadmin，输入 cluster heartbeat interval n 将时钟同步的间隔时间设置成 n 秒钟。

如果用户的应用的 CPU 负载不会达到 100%，或者是 100%CPU 负载持续时间在 15（3*5）秒以内，则不需要改动参数。

9) **Q:** GreatTurbo Cluster Server 10 是否能够检测网卡的硬件连接情况？

**A:** 能够。GreatTurbo Cluster Server 10 通过配置，能够检测服务所使用的网卡是否正常连接到交换机或者 HUB 上。

**注意：待检测的网卡的网线另一段必须是连接在交换机或者 HUB 上。**

10) **Q:** GreatTurbo Cluster Server 10 服务迁移时间是否可以进行调整？

**A:** 可以。通过调整服务检测到错误的时间来调整服务迁移时间。由于服务检测到错误的时间＝服务检测的间隔时间 * 服务连续出错的次数，所以在配置服务的检测脚本时，通过调整 check interval 以及 check count 参数，可以控制服务检测到错误的时间。

通常，习惯于将 check count（单位为次）设置为 3，表示连续 3 次出错后才确认服务的确出错，以保证检测的可靠性。而 check interval（单位为秒）的设置由用户的应用决定。

一定要根据用户的应用要求来合理的设置 check interval 的值。

如果应用更着重于稳定性/可靠性（属于电信/银行等关键业务的应用或应用本身的负载较重等），建议将 interval 的值适当调大，如设置 check interval 在 10 秒-30 秒之间，当服务出错，服务最少将在 30 秒-90 秒之后进行服务的迁移。

如果应用更着重于性能（应用负载较轻或者应用要求极短的服务不可用时间），建议将 interval 的值适当调小，如设置 check interval 在 3 秒-10 秒之间，当服务出错，服务最少将在 9 秒-30 秒之后进行服务的迁移。

对于通常的应用，一般将 check interval 设置成 5 秒，当服务出错，服务最少将在 15 秒之后进行服务的迁移。

11) **Q:**当服务出现 error 状态时该如何处理？

**A:** 当服务启动或停止失败时，服务将处在错误状态，处在错误状态下的服务，其资源的状态将无法确定（例如，它的部分资源已经释放，但某些服务资源仍然将在所有者节点上进行配置）。GreatTurbo Cluster Server 10 中，以下的情形将导致服务的状态为 error：①启用服务失败；②停止服务失败(多为服务检测失败时停止服务)；③禁用服务失败；④切换服务失败；⑤停止 GreatTurbo Cluster Server 10 失败(即/etc/init.d/cluster stop 失败)。

由于处理错误状态下的服务可能有资源没有释放，所以处理错误状态中的服务必须十分小心。如果服务资源仍然在所有者节点上配置，那么在另一个集群节点上启动服务可能会造成严重的问题。例如，如果某一文件系统已经加载到它的所有者节点上，您又在另一个集群节点上启动该服务，则该文件系统将可能同时在两个系统上被加载，这样可能导致数据损坏。所以，您必须在确认所有的服务资源都已经完全释放掉了之后才能在另一个集群节点上重新启用服务。

对于 errror 状态服务，若按服务配置的重启策略自动重启后仍然不成功，就需要手动干预。手动干预处理错误状态的服务的过程一般如下：

①定位错误的原因。当服务的状态变为 error 时，首先要定位错误的原因，要从 GreatTurbo Cluster Server 10 的日志中分析。一般服务启动和停止的过程分为几步(例如，启动服务时会有 mount 分区，启动活动 IP，执行启动脚本等几步)，每一步出错都会打印日志，要从日志中分析具体是哪一步出错了。

②排除故障。当找到错误的原因后，需要排除故障。一般是通过先 modify，使服务配置正确，然后 disable 操作，使服务能够成为 disabled 状态，若操作成功就可以进行 enable 或 delete 等其它的操作了。在 modify 的过程中，建议修改服务配置使 Reboot After Fail 为 no，以避免在 modify 服务仍有问题的时候进行无谓的重启。例如，如果是启动/停止脚本写的有错误，则需要纠正错误并测试正确等。

③释放资源，恢复系统。故障排除后，需要释放所有的没有释放的服务资源，并重新恢复系统。

12) **Q:** 使用磁盘镜像设备(drbd)时，如何配置 Oracle 的服务？

**A:** 当使用磁盘镜像设备时，如果想配置 Oracle 的服务，建议将 Oracle 在 GreatTurbo Cluster Server 10 的两个节点上分别安装，即不安装在磁盘镜像设备上，而安装在节点自己的磁盘设备上，安装时先不要建库，在安装完后再单独建库。将 Oracle 的库文件建在磁盘镜像设备上，建库时，先在一个节点上 mount 磁盘镜像设备、建库，完成后，再在另一个节点上 mount 磁盘镜像设备、以和前一个节点上完全相同的操作建库，并选择覆盖原来的库。这样，Oracle 就在磁盘镜像设备上安装完成了，然后先手动试一下能否启动 Oracle，如果手动能够启动 Oracle，就可以在 GreatTurbo Cluster Server 10 中配置基于磁盘镜像设备的 Oracle 服务了。

## 9.4 诊断和纠正集群中的故障

为了确保您能识别集群中的问题，您必须启用事件日志记录。此外，一旦您发现集群中有问题，一定要设置严重性等级，以对集群 daemon 进行调试。这样会记录下说明信息，以帮助您解决问题。

如果您在运行 cluadmin 工具时出现故障（比如：无法启用某一服务），您可以对 svcmgr daemon 设置严重性等级，以便进行调试。这样当您运行 cluadmin 工具时，就会显示调试信息。更多详情，请参见**修改集群事件日志**。

请使用下表诊断和纠正集群中的故障。

| 故障 | 症状 | 解决方案 |
|---|---|---|
| SCSI 总线未终止 | 在日志文件中出现 SCSI 错误 | 在总线起始处，必须终止每一条 SCSI 总线。根据总线配置的不同，您可能需要启用或禁用主机总线适配器、RAID 控制器和存储附件中的终止功能。如 |

| 故障 | 症状 | 解决方案 |
| --- | --- | --- |
| | | 果您希望支持热插拔功能，则您必须使用外部终止功能来终止 SCSI 总线。<br><br>此外，还需确保不会有设备使用超过 0.1 米的短线（stub）连接到 SCSI 总线。<br><br>有关终止不同类型 SCSI 总线的信息，请参见配置共享磁盘存储器**和 SCSI 总线终止**。 |
| SCSI 总线长度超过最大限制 | 在日志文件中出现 SCSI 错误 | 每种类型的 SCSI 总线必须符合 SCSI 总线长度中所规定的长度限制。<br><br>此外，还要确保没有单端设备连接到 LVD SCSI 总线上，因为这将导致全部总线返回到某一条单端总线（比差分总线有更严格的长度限制）上。 |
| SCSI 总线标识号并非唯一 | 在日志文件中出现 SCSI 错误 | SCSI 总线上的每台设备必须有一个唯一的标识号。如果您有一条多启动程序 SCSI 总线，就必须为连接到该总线上的其中一台主机总线适配器修改缺省的 SCSI 标识号（7），并确保所有的磁盘设备都有一个唯一的标识号。更多详情，请参见 **SCSI 标识号**。 |
| SCSI 命令在完成之前超时 | 在日志文件中出现 SCSI 错误 | 某一 SCSI 总线上的优先级仲裁方案（arbitration scheme）可能会导致低优先级设备被锁定一段时间。如果某台低优先级存储设备（如磁盘等）不能获得判优并完成主机排队分配给它的命令，就可能导致超时。对于某些工作负载，您可以通过将低优先级 SCSI 标识号分配给主机总线适配器来避免这一问题。更多详情，请参见 **SCSI 标识号**。 |
| 服务文件系统不干 | 某一被禁用的服 | 人工运行某个检查程序（如 fsck）， 然后启用该服 |

| 故障 | 症状 | 解决方案 |
| --- | --- | --- |
| 净 | 务无法启用。 | 务。<br><br>注意，集群基础设置不会自动修复文件系统的不一致性（比如使用 fsck-y 命令）。这要求集群管理员在纠正过程中进行干预，并发现损坏文件和受影响的文件。 |
| 集群服务操作失败 | 控制台上或日志文件中出现操作失败的信息。 | 服务操作失败的原因有很多种（比如服务停止或启动）。为了帮助您识别问题的原因，您可以设置严重性等级对集群 daemon 进行调试，从而记录说明信息。然后，重新尝试操作并检查日志文件。更多详情，请参见**修改集群事件日志**。 |
| 由于某一文件系统无法卸载而导致集群服务停止失败 | 控制台上或日志文件中出现操作失败的信息。 | 使用 fuser 和 ps 命令识别正在访问文件系统的程序，使用 kill 命令停止该程序。您也可以使用 lsof -t file_system 命令显示正在访问指定文件系统的程序的标识号， 或者将输出结果传输给 kill 命令。<br><br>为避免出现此问题，您必须确保只有与集群相关的程序才能访问共享存储数据。此外，您可能希望修改服务并对文件系统启用强制卸载。这样即使集群服务被某个应用或用户访问时也可卸载文件系统。 |
| 集群数据库中条目错误 | 集群服务受损 | 检查每个集群系统上的/etc/opt/cluster.cluster.conf 文件。如果文件中的某个条目错误，可根据**修改集群配置**中的说明，通过运行 member_config 工具来修改集群配置并纠正问题。 |
| 集群数据库或 /etc/hosts 文件中以太网心跳条目错 | 集群状态显示以太网心跳通道 OFFLINE（离线） | 检查每个集群系统上的/etc/opt/cluster/cluster.conf 文件，并确认为 chan0 所指定的网络接口名称与集群系统上 hostname 命令所返回的名称一致。如果 |

| 故障 | 症状 | 解决方案 |
| --- | --- | --- |
| 误 | （尽管接口仍然有效）。 | 文件中的某个条目错误，可根据**修改集群配置**中的说明，通过运行 member_config 工具来修改集群配置并纠正问题。<br><br>如果 cluster.conf 文件中的条目正确，则检查 /etc/hosts 文件并确保其包含所有的网络接口条目。同时，要确保/etc/host 文件使用正确的格式。更多详情，请参见**编辑/etc./hosts 文件**。<br><br>此外，您必须确保能使用 ping 命令将某个数据包发送到集群中使用的所有网络接口上。 |
| 心跳通道故障 | 心跳通道状态为 OFFLINE（离线） | 检查每个集群系统上的/etc/opt/cluster/cluster.conf 文件，并确认为每个串行心跳通道所指定的设备专用文件都与连接通道的实际串行口相匹配。如果文件中的某个条目错误，可根据**修改集群配置**中的说明，通过运行 member_config 工具来修改集群配置并纠正问题。<br><br>确认每一个心跳通道都使用正确的线缆类型进行了连接。<br><br>确认您能通过网络接口为每条以太网心跳通道“ping”通每个集群系统。 |

## 9.5 联系拓林思软件有限公司

1) 邮件联系：support@turbolinux.com.cn。

请在邮件中详细描述 GreatTurbo Cluster Server 10 的版本信息，故障现象，并在附件中附上用 log 收集工具收集的 log 信息。

2) 电话联系：01065054020

# 附录 A　补充硬件信息

以下信息可帮助您设置集群硬件配置。在某些情况下，这些信息因厂商而异。

- 设置 Cyclades 终端服务器
- 设置 RPS-10 电子开关
- SCSI 总线配置要求

## A.1 设置 Cyclades 终端服务器

本文提供了有关设置 Cyclades 终端服务器的信息，可帮助您设置终端服务器。

Cyclades 终端服务器包括 2 个主要部分：

- PR3000 路由器

  该路由器通过一根传统网络线缆连接到网络交换机上（或直接连接在网络上）。

- 异步串行扩展器

  该模块连接在 PR3000 路由器上，可提供 16 个串行口。虽然您可以连接 4 个模块，但如果想获得最佳可靠性，则只能连接 2 个模块。您可以使用 RJ45 到 DB9 交叉线缆将每个系统连接到串行扩展器上。

如欲设置 Cyclades 终端服务器，请遵循以下步骤进行操作：

- 为路由器设置一个 IP 地址。
- 配置网络参数和终端端口参数。
- 配置 Turbolinux，以向控制台端口发送控制台信息。
- 连接到控制台端口。

### A.1.1 设置路由器的 IP 地址

设置 Cyclades 终端服务器的第一步是为 PR3000 路由器指定一个互联网协议（IP）地址。请遵循以下步骤进行：

- 使用 RJ45 到 DB9 交叉线缆将路由器的串行控制台端口连接到某个系统上的串行口。
- 根据控制台登录提示[PR3000]，使用 Cyclades 手册提供的密码登录到超级账户中。
- 控制台将显示一系列菜单。 请按顺序选择以下菜单项：Config（配置）、Interface（接口）、Ethernet（以太网）和 Network Protocol（网络协议）。然后输入 IP 地址和其它信息。例如：

```
Cyclades-PR3000（PR3000）Main Menu


1. Config                   2. Applications        3. Logout
4. Debug                     5. Info                  5. Admin


Select option ==> 1


Cyclades-PR3000（PR3000）Config Menu


1. Interface                2. Static Routes       3. System
4. Security                  5. Multilink           6. IP
7. Transparent Bridge       8. Rules List          9. Controller


（L for list）Select option ==> 1


Cyclades-PR3000（PR3000）Interface Menu
```

```
1. Ethernet                    2. Slot 1（Zbus-A）

（L for list）Select option ==> 1

Cyclades-PR3000（PR3000）Ethernet Interface Menu

1. Encapsulation               2. Network Protocol     3. Routing Protocol
4. Traffic Control

（L for list）Select option ==> 2

（A）ctive or（I）nactive [A]：
Interface（U）nnumbered or（N）umbered [N]：
Primary IP address： 111.222.3.26
Subnet Mask [255.255.255.0]：
Secondary IP address [0.0.0.0]：
IP MTU [1500]：
NAT - Address Scope（（L）ocal,（G）lobal, or Global（A）ssigned）[G]：
ICMP Port（（A）ctive or（I）nactive）[I]：
Incoming Rule List Name（? for help）[None]：
Outgoing Rule List Name（? for help）[None]：
Proxy ARP（（A）ctive or（I）nactive）[I]：
IP Bridge（（A）ctive or（I）nactive）[I]：

ESC

（D）iscard, save to（F）lash or save to（R）un configuration：F

Changes were saved in Flash configuration !
```

## A.1.2 设置网络和终端端口参数

在为 PR3000 路由器指定一个 IP 地址后，您必须设置网络和终端端口参数。

根据控制台登录提示[PR3000]，使用 Cyclades 手册提供的密码登录到超级账户中。 控制台将显示一系列菜单。 输入适当的信息。例如：

```
Cyclades-PR3000（PR3000）Main Menu

1. Config                2. Applications        3. Logout
4. Debug                 5. Info                5. Admin

Select option ==> 1

Cyclades-PR3000（PR3000）Config Menu

1. Interface             2. Static Routes       3. System
4. Security              5. Multilink           6. IP
7. Transparent Bridge    8. Rules List          9. Controller

（L for list）Select option ==> 1

Cyclades-PR3000（PR3000）Interface Menu

1. Ethernet              2. Slot 1（Zbus-A）

（L for list）Select option ==> 1
```

```
Cyclades-PR3000（PR3000）Ethernet Interface Menu

1. Encapsulation          2. Network Protocol    3. Routing Protocol
4. Traffic Control

（L for list）Select option ==> 1

Ethernet（A）ctive or（I）nactive [A]：
MAC address [00：60：2G：00：08：3B]：

Cyclades-PR3000（PR3000）Ethernet Interface Menu

1. Encapsulation          2. Network Protocol    3. Routing Protocol
4. Traffic Control

（L for list）Select option ==> 2

Ethernet（A）ctive or（I）nactive [A]：
Interface（U）nnumbered or（N）umbered [N]：
Primary IP address [111.222.3.26]：
Subnet Mask [255.255.255.0]：
Secondary IP address [0.0.0.0]：
IP MTU [1500]：
NAT - Address Scope（（L）ocal,（G）lobal, or Global（A）ssigned）[G]：
ICMP Port（（A）ctive or（I）nactive）[I]：
Incoming Rule List Name（? for help）[None]：
Outgoing Rule List Name（? for help）[None]：
Proxy ARP（（A）ctive or（I）nactive）[I]：
IP Bridge（（A）ctive or（I）nactive）[I]：
```

```
Cyclades-PR3000（PR3000）Ethernet Interface Menu

1. Encapsulation          2. Network Protocol    3. Routing Protocol
4. Traffic Control

（L for list）Select option ==>

Cyclades-PR3000（PR3000）Interface Menu

1. Ethernet               2. Slot 1（Zbus-A）

（L for list）Select option ==> 2

Cyclades-PR3000（PR3000）Slot 1（Zbus-A）Range Menu

1. ZBUS Card               2. One Port              3. Range
4. All Ports

（L for list）Select option ==> 4

Cyclades-PR3000（PR3000）Slot 1（Zbus-A）Interface Menu

1. Encapsulation          2. Network Protocol    3. Routing Protocol
4. Physical               5. Traffic Control     6. Authentication
7. Wizards

（L for list）Select option ==> 1
```

```
Cyclades-PR3000（PR3000）Slot 1（Zbus-A）Encapsulation Menu

1. PPP                    2. PPPCHAR              3. CHAR
4. Slip                   5. SlipCHAR             6. Inactive

Select Option ==> 3

Device Type（（T）erminal,（P）rinter or（S）ocket ）[S]：
TCP KeepAlive time in minutes（0 - no KeepAlive, 1 to 120）[0]：
（W）ait for or（S）tart a connection [W]：
Filter NULL char after CR char（Y/N）[N]：
Idle timeout in minutes（0 - no timeout, 1 to 120）[0]：
DTR ON only if socket connection established（（Y）es or（N）o ）[Y]：
Device attached to this port will send ECHO（Y/N）[Y]：

Cyclades-PR3000（PR3000）Slot 1（Zbus-A）Encapsulation Menu

1. PPP                    2. PPPCHAR              3. CHAR
4. Slip                   5. SlipCHAR             6. Inactive

Select Option ==>

Cyclades-PR3000（PR3000）Slot 1（Zbus-A）Interface Menu

1. Encapsulation          2. Network Protocol     3. Routing Protocol
4. Physical               5. Traffic Control      6. Authentication
7. Wizards

（L for list）Select option ==> 2
```

```
Interface IP address for a Remote Telnet [0.0.0.0]：

Cyclades-PR3000（PR3000）Slot 1（Zbus-A）Interface Menu

1. Encapsulation            2. Network Protocol    3. Routing Protocol
4. Physical                 5. Traffic Control     6. Authentication
7. Wizards

（L for list）Select option ==> 4

Speed（? for help）[115.2k]：9.6k
Parity（（O）DD,（E）VEN or（N）ONE ）[N]：
Character size（ 5 to 8 ）[8]：
Stop bits（1 or 2 ）[1]：
Flow control（（S）oftware,（H）ardware or（N）one ）[N]：
Modem connection（Y/N）[N]：
RTS mode（（N）ormal Flow Control or（L）egacy Half Duplex ）[N]：
Input Signal DCD on（ Y/N ）[N]：n
Input Signal DSR on（ Y/N ）[N]：
Input Signal CTS on（ Y/N ）[N]：

Cyclades-PR3000（PR3000）Slot 1（Zbus-A）Interface Menu

1. Encapsulation            2. Network Protocol    3. Routing Protocol
4. Physical                 5. Traffic Control     6. Authentication
7. Wizards

（L for list）Select option ==> 6
```

```
Authentication Type（（N）one,（L）ocal or（S）erver ）[N]：

ESC

（D）iscard, save to（F）lash or save to（R）un configuration：F

Changes were saved in Flash configuration
```

### A.1.3 配置 Turbolinux，以向控制台端口发送控制台信息

设置完网络和终端端口参数后，您可以配置 Linux，以向控制台串行口发送控制台信息。按照要求在每个集群系统上进行如下操作：

1. 确保该集群系统为串行控制台输出而配置。通常在缺省情况下启用此项支持功能。您必须设置下列内核选项：

   ```
   CONFIG_VT=y
   CONFIG_VT_CONSOLE=y
   CONFIG_SERIAL=y
   CONFIG_SERIAL_CONSOLE=y
   ```

2. 当指定内核选项后，在 Character Devices（字符设备）项下，选择 Support for console on serial port（支持控制台串行口）。
3. 编辑/etc/lilo.conf 文件。将下面的程序行添加到文件的顶端条目中，以指定系统将串行口用作控制台：

   serial=0,9600n8

针对每一个可引导的内核程序，将与下列相似的程序行添加到 stanza 条目中，从而把内核信息发送到指定的控制台串行口（例如 ttyS0）和图形终端上：

append="console=ttyS0 console=tty1"

下面是一个/ect/lilo.conf 文件的举例：

```
boot=/dev/hda
map=/boot/map
install=/boot/boot.b
prompt
timeout=50
default=scons
serial=0,9600n8

image=/boot/vmlinuz-2.2.12-20
label=linux
initrd=/boot/initrd-2.2.12-20.img
read-only
root=/dev/hda1
append="mem=127M"

image=/boot/vmlinuz-2.2.12-20
label=scons
initrd=/boot/initrd-2.2.12-20.img
read-only
root=/dev/hda1
append="mem=127M console=ttyS0 console=tty1"
```

4. 通过调用/sbin/lilo 命令对/etc/lilo.conf 文件进行更改。
5. 如果想通过控制台串行口（例如 ttyS0）启用登录，则编辑/etc/inittab 文件（其中包含 getty 定义），并添加与下列相似的程序行：S0：2345：respawn：/sbin/getty ttyS0 DT9600 vt100
6. 通过在/etc/securetty 文件中的某一程序行上指定串行口使根目录能登录到串行口上。例如：

   ttyS0

7. 重新创建/dev/console 设备专用文件，以便为串行口指定主号码。例如：

```
# ls -l /dev/console
crw--w--w-    1 joe    root    5,    1 Feb 11 10：05 /dev/console
# mv /dev/console /dev/console.old
# ls -l /dev/ttyS0
crw-------    1 joe    tty    4,   64 Feb 14 13：14 /dev/ttyS0
# mknod console c 4 64
```

### A.1.4 连接到控制台端口

如果想连接到控制台端口，则使用以下 telnet 命令格式：

telnet hostname_or_IP_address port_number

指定与终端服务器串行程序行有关的集群系统的主机名称或其 IP 地址以及端口号。端口号的变化范围为 1 到 16，您可以通过将端口号添加到 31000 来指定它。比如：您可以指定从 31001 到 31016 之间的某个端口号。

下面的举例将 cluconsole 系统连接到了端口 1 上：

# telnet cluconsole 31001

下面的举例将 cluconsole 系统连接到了端口 16 上：

# telnet cluconsole 31016

下面的举例将带有 IP 地址 111.222.3.26 的系统连接到了端口 2 上：

# telnet 11.222.3.26 31002

登录以后，您输入的任何内容都会被重复。例如：

```
[root@localhost /root]# date
date
Sat Feb 12 00：01：35 EST 2000
[root@localhost /root]#
```

要想纠正这种情况，就必须转换操作模式，这种操作模式已经经过 telnet 和终端服务器协商。下面的例子使用了^}换码符：

```
[root@localhost /root]# ^]
telnet> mode character
```

在本地目录下，您可以通过创建.telnetrc 文件发出模式符命令，如下所示：

```
cluconsole
mode character
```

## A.2 设置 RPS-10 电子开关

如果您在集群系统中使用了 RPS-10 系列电子开关，您就必须：

- 在两个电子开关上把循环地址设置为 0，并确保开关的位置正确，没有悬在中间。
- 在将两个电子开关上上的四个 SetUp（设置）按钮，见下表：

**表 1.**

| 开关 | 功能 | 上位 | 下位 |
|---|---|---|---|
| 1 | 数据速率 | ? | X |
| 2 | 切换延迟 | ? | X |
| 3 | 加电缺省 | X | ? |
| 4 | 空 | ? | X |

- 确保在/etc/opt/cluster/cluster.conf 文件中指定的串行口设备专用文件（如， /dev/ttyS1）与电子开关的串行线缆连接的串行口相对应。
- 将每个集群系统的电源线连接到它的电子开关上。
- 用空调制解调器线缆连接到为其它集群系统供电的电子开关上，从而连接每个集群系统的串行口。

下图显示了 RPS-10 系列电子开关的配置。

RPS-10 电子开关硬件配置

如想获取额外的安装信息，请查看厂商提供的 RPS-10 文件。注意，厂商信息以本文提供的信息为准。

# A.3 SCSI 总线配置要求

为确保正确操作，SCSI 总线必须符合各项配置要求。如不遵守这些要求，集群操作、集群应用和数据的实用性就会受到影响。

您必须遵守以下 SCSI 总线配置要求：

- 总线必须在每个末端终止。而且，您如何终止 SCSI 总线将会影响到您是否可以使用热插拔功能。更多详情，请参见 SCSI 总线终止。
- TERMPWR（终结器电源）必须由连接到总线的主机总线适配器提供。更多详情，请参见 SCSI 总线终止。
- 主动式 SCSI 终结器只能在一个多启动程序总线中使用。更多详情，请参见 SCSI 总线终止。
- 对于每种总线类型，其扩展的长度都不能超过其最大限制。 SCS 总线长度包括内部线缆的长度。更多详情，请参见 SCSI 总线长度。
- 总线上的所有设备（主机总线适配器和磁盘）都必须有唯一的 SCSI 标识号。查看 SCSI 标识号可获得更多信息。
- 对于每个共享的 SCSI 设备，其 Linux 名称必须和集群系统上的名称一致。比如：对于在一个集群系统中名称为/dev/sdc 的设备，在其它集群系统中也要以/dev/sdc 命名。通常，您可以在两个集群系统中使用同样的硬件，来确保设备名称一致。
- 如果您使用了 SCSI 保留，您就必须为主机总线适配器启用总线复位。最好是启用总线复位，因为如果不被启用它，总线适配器驱动程序将不能正确运行。最新的 Turbolinux Servers 包括 Linux 内核，它能正确处理 SCSI 总线复位

在设置 SCSI 标识号时，需要禁用主机总线适配器终止和总线复位，并且应使用系统配置工具。当系统启动时，将会显示出如何运行配置工具的信息。比如,该信息可能会指导您按下 Ctrl-A, 并按照提示执行随后的任务。如果想设置存储附件和 RAID 控制器终止，请查看厂商提供的文件。更多详情，请参见 SCSI 总线终止和 SCSI  标识号。

关于 SCSI 总线要求的详细信息，请访问 www.scsita.org 和阅读以下章节。

### A.3.1 SCSI 总线终端

SCSI 总线是两台终结器之间的电路。每一台设备（主机总线适配器、RAID 控制器，或磁盘等）都通过一条短线连接到 SCSI 总线上，短线长度通常不超过 0.1 米，是一条未终止的总线线段。

在总线两端必须只放置两个终结器。终结器过多、终结器不在总线两端或短线较长都将导致总线错误运行。如果内部（板载）设备终端可以被禁用，则可以通过连接到该总线上的设备或外部终结器来提供某一 SCSI 总线的终端。

终结器通过一条 SCSI 配电线路（或信号）TERMPWR 提供电源，这样只要总线上有一台供电器，终结器就可以工作。在某一集群中，TERMPWR 必须由主机总线适配器（而非附件中的磁盘）提供。通常，您可以通过在驱动器上设置一根跳线来禁用某个磁盘中的 TERMPWR。更多详情，请参见**磁盘驱动器文档**。

另外，还有两种类型的 SCSI 终结器。主动式终结器和被动式终结器，前者为 TERMPWR 提供了一个稳压器，而后者则在 TERMPWR 和地面之间提供了一个电阻网络。但被动式终结器很容易受到 TERMPWR 波动的影响，因此建议您在在集群中使用主动式终结器。

为了便于维护，需要求存储配置可支持热插拔功能（也就是说，在维护总线终端和操作过程中，可以将主机总线适配器从某一 SCSI 总线断开）。但如果您使用的是单启始端 SCSI 总线，就不必使用热插拔了，因为当您拆除某台主机时，专用总线无需保持运行。关于热插拔配置的举例，请参见**设置多启始端 SCSI 总线配置**。

如果您使用的是多启始端 SCSI 总线，那么您必须遵守以下热插拔要求：

- SCSI 设备、终结器和线缆必须符合严格的热插拔要求（请参见附录 D SCSI 并行接口-3（SPI-3）中介绍的最新 SCSI 规范）。您可以访问 www.t10.org 来获得该文件。

- 必须禁用内部主机总线适配器终端。并非所有的适配器都支持这一特性。
- 如果某台主机总线适配器位于 SCSI 总线的末端，则外部终结器必须提供总线终端。
- 用于将主机总线适配器连接到 SCSI 总线的短线长度必须少于 0.1 米。在系统附件内部使用一条长线缆来连接隔板（bulkhead）的主机总线适配器不能支持热插拔功能。此外，有内部连接器和线缆（扩展系统附件内部的总线）的主机总线适配器也不能支持热插拔功能。注意，任何内部线缆都必须包括在 SCSI 总线长度内。

当您从支持热插拔功能的某一单启始端 SCSI 总线或多启始端 SCSI 总线断开某台设备时，需要遵循以下操作规范：

- 禁止将未终止的 SCSI 线缆连接到某台运行的主机适配器或存储设备上。
- 当 SCSI 线缆被断开时，接线插脚不得弯向或接触电导体。
- 如果您想从单启始端总线上断开主机总线适配器，必须先从 RAID 控制器上断开 SCSI 线缆，然后再断开适配器。 这样可确保不会对 RAID 控制器执行任何暴露错误的输入。
- 您需要戴上接地防静电护腕，通过物理方式保护线缆末端不与其它物体接触来断开 SCSI 线缆，同时必须防止插头产生静电放电。
- 不得拆除当前正参与 SCSI 总线处理活动的任何设备。.

如欲启用或禁用某个适配器的内部终端，请使用系统 BIOS 工具。 当系统启动时，将显示一条说明如何启动该程序的信息。 例如：信息可能指示您按下 Ctrl-A，并根据提示设置终端。 此时，您也可以根据需要设置 SCSI 标识号和禁用 SCSI 总线重置。 更多详情，请参见 SCSI 标识号。

如欲设置存储附件和 RAID 控制器终端，请参见厂商文档。

## A.3.2 SCSI 总线长度

SCSI 总线必须符合总线类型的长度限制。不符合这些限制的总线将无法正确操作。一条总线的长度应以从一个终端到另一个终端的长度来计算，并且必须包括存在于系统或存储附件中的线缆长度。

集群支持 LVD（低压差分）总线。一条单启始端 LVD 总线的最大长度为 25 米。一条多启始端 LVD 总线的最大长度为 12 米。根据 SCSI 标准，一条单启始端 LVD 总线就是一条只连接两台设备（每台设备距离其中一台终结器 0.1 米）的总线，而其它所有总线均被定义为多启始端总线。

禁止将任何单端设备连接到某条 LVD 总线上，或者将该总线转换到某条单端总线上，因为单端总线比微分总线的最大长度短得多。

## A.3.3 SCSI 标识号

SCSI 总线上的设备都必须有一个唯一的 SCSI 标识号。这些设备包括主机总线适配器、RAID 控制器和磁盘。

SCSI 总线上设备的号码取决于总线的数据通道。集群系统支持宽 SCSI 总线，它是 16 位的数据通道，最多支持 16 台设备。因此，共有 16 个 SCSI 标识号，供您分配给总线上的设备。

而且，SCSI 标识号划分了优先次序。使用下面的优先次序来分配 SCSI 标识号：

7 - 6 - 5 - 4 - 3 - 2 - 1 - 0 - 15 - 14 - 13 - 12 - 11 - 10 - 9 - 8

从上面的优先次序中可以看出 7 拥有最高的优先级,8 的优先级优先级最低。7 是分给主机总线适配器的缺省 SCSI 标识号 ，因为通常适配器都拥有最高的优先级。在多启始端总线上，一定要更改其中一个主机总线适配器的 SCSI 标识号，以防止出现完全相同的数值。

可以手动（在磁盘上设置跳线）和自动（根据附件插槽号码）为 JBOD 附件中的磁盘分配 SCSI 标识号。通过使用 RAID 管理界面，您还可以在 RAID 子系统中为逻辑单元分配标识号。

修改适配器的 SCSI 标识号需要使用系统 BIOS 工具。系统重启时，将会显示如何启动该工具的信息显示。比如：您可能会被要求按下 Ctrl-A，然后根据提示设置 SCSI 标识号。此时，您也可以根据需要对适配器的内部终止和禁止 SCSI 总线复位进行设置。更多详情，请参见 SCSI 总线中止。

SCSI 总线上的优先仲裁方案可能会使低优先级的设备被锁定一段时间。这样会导致命令超时，如果一个低优先级存储设备（如磁盘）不能被判优并完成主机已经分配给它的命令，就会导致超时。对于一些工作负载，您可以通过把低优先级设备的 SCSI 标识号分配给主机总线适配器来避免这个问题。

# 附录 B　补充软件信息

以下信息能帮您管理集群软件配置：

- 集群通信机制
- 集群守护进程（Daemons）
- 故障切换和恢复情形
- 集群数据库字段
- 调整 Oracle 服务
- 在 Turbolinux Cluster Server 上使用 GreatTurbo Cluster Server 10

## B.1 集群通信机制

通过使用几种集群间通信机制，集群系统可在发生错误时保证数据的完整性，并更改集群的行为。集群采用的机制包括：

- 当系统成为集群成员时对其加以控制。
- 确定集群系统的状态。
- 当发生错误时，对集群的行为进行控制。

集群通信机制包括：

- 远程电子开关监控

  每个集群系统定期监控远程电子开关（如果有的话）连接线路的状态。集群系统可以使用这种信息来确定其它集群系统的状况。如果电子开关通信机制完全失效，将不能自动导致故障切换。

- 以太网和串行心跳

集群系统使用以太网（点对点或借助集线器）和串行线连接在一起。每个集群系统都会定期通过这些线路发出心跳（ping）。集群系统将使用这些信息来确定系统的状况，以保证正确的集群操作。如果心跳通信机制完全失效，将会自动导致故障切换。

## B.2 集群守护进程（Daemon）

集群守护进程包括：

- syncd 守护进程

  在每个集群系统中，syncd 守护进程可以收到对方节点关于本地节点状态标记的请求，并将结果发送给它。

- 心跳守护进程

  在每个集群系统中，hb 心跳守护进程通过点对点以太网和两个集群系统连接的串行线发出 ping

- 电源守护进程

  在每个集群系统中，powerd 电源守护进程负责对远程电子开关连接进行监控。

- 服务管理器守护进程

  在每个集群系统中，svcmgr 服务管理器守护进程负责通过停止和启动服务的方法来响应集群成员的变化。

- 服务检查守护进程

  在每个集群系统中，scvcheck 服务检查管理器守护进程负责定期执行**服务应用代理**来检查服务的状态。如果应用代理返回了一个故障，服务将不能正常进行，故障切换就会被触发。

- 集群监控守护进程

  在每个集群系统中，集群 监控守护进程均通过 inittab 来启动，并且当出现 inittab 时，它会立即做出响应。它会监控其它守护进程，以确保它们在正常运行。如果某些守护进程突然退出，集群监控守护进程将立即对此进行响应。

## B.3 故障切换和恢复情形

充分了解发生重要事件时集群的行为可以帮助您管理集群系统。注意，集群行为取决于您在配置中是否使用电子开关。

下面描述了系统对各种故障和错误情形的响应：

- 系统挂起
- 系统紧急（Panic）
- 网络连接完全故障
- 远程电子开关连接故障
- 集群守护进程故障

### B.3.1 系统挂起（Hang）

在使用电子开关的集群配置中，如果系统 " 挂起 "，集群系统会做出如下行为：

1. 正常运行的集群系统检测到 " hung " （挂起）的集群系统不能通过心跳通道进行通信。
2. 正常运行的集群系统通过电子开关、sg 保留或软件重启动，对 " hung " （挂起）的集群系统重启动。
3. 正常运行的集群系统将重新运行曾经在 " hung " （挂起）的系统上运行的任何一项服务。

4. 如果先前的“ hung ”（挂起）系统重新启动， GreatTurbo HA 将会对网络连接进行全面检查。一旦恢复网络连接，系统将重新加入集群；随后，服务将会根据每项服务的放置政策得到重新平衡。

### B.3.2 系统紧急（Panic）

系统紧急是对检查到的软件错误进行的控制响应。系统 panic 可通过关闭系统来确保将其返回到一致状态。一旦出现集群系统，将会导致下列情况：

1. 正常运行的集群系统检测到集群系统不能通过心跳通道进行通信。
2. 发生紧急的集群系统会触发系统关机和重启
3. 当您使用电子开关时，正常运行的集群系统可以为发生紧急的集群系统重启电源。
4. 正常运行的集群系统可以重新启动曾在出现 panic 的系统上运行的任何一项服务。
5. 系统重新启动时，GreatTurbo Cluster Server10 对网络连接进行全面检查。一旦恢复网络连接，系统将重新加入集群；随后，服务将会根据每项服务的放置政策得到重新平衡。

### B.3.4 网络连接完全故障

当系统之间的心跳网络连接失败时，就会出现网络连接完全故障。这可能是由以下原因导致：

- 所有的心跳网络线缆与系统断开。
- 用于心跳通信的所有串行连接和网络接口发生了故障。

一旦出现网络连接完全故障，正在检测故障的系统就会自己重新启动，并将运行的所有服务重新定位到对方节点。

如果系统重启，GreatTurbo Cluster Server 10 将对网络连接进行全面检查。一旦恢复网络连接，系统将重新加入集群；随后，服务将会根据每项服务的放置政策得到重新平衡

### B.3.5 远程电子开关连接故障

当您向远程电子开关发出的查询请求出现故障，而两个系统仍在运行时，集群行为不会发生改变，除非集群系统所连接的电源开关正好是对其它系统重启电源的远程电子开关。电源守护进程将持续记录拥有高优先级的信息，如电子开关故障或电子开关中断连接（比如线缆被断开）。

如果集群系统试图使用一个发生故障的远程电子开关，那么在发生故障的系统上运行的服务就会被终止。 但为了确保数据完整性，它们不会被故障切换至其它集群系统中。相反，它们会被停止，直到硬件故障被排除为止。

### B.3.6 集群 Daemon 故障

如果集群守护进程（syncd、svcmgr、svccheck、powerd、hb）在集群系统上发生故障，它将被集群监控守护进程重启动（respawned）。而且，如果集群监控守护进程发生故障，它将被 init 进程重启动（respawned）。

## B.4 集群数据库选项

集群数据库的备份存在/etc/opt/cluster/cluster.conf 文件中。备份内容包括集群成员和集群服务的详细信息。请不要手动编辑配置文件，而应使用集群程序来修改集群配置。

当您运行 member_config 脚本时，您指定的站点信息将被加入一些字段，这些字段位于数据库中的[members]（成员）部分。下面是对集群成员字段的描述：

**表 1.**

| | |
|---|---|
| start member0 start chan0 device = serial_port type = serial end chan0 | 为串行心跳通道指定一个 tty 端口，这个端口将连接到空调制解调器线缆（null model cable）上。比如：serial_port（串行口）可以是/dev/ttyS1。 |
| start chan1 name = interface_name type = net end chan1 | 为以太网心跳通道指定一个网络接口。interface_name 是指将接口分配给某台主机时，该主机的名称（如：storage0）。 |
| start chan2 device = interface_name type = net end chan2 | 给网络接口指定第二条以太网心跳通道。interface_name 是指将接口分配给某台主机时，该主机的名称（如：storage0）。本字段可用以指定点对点专用心跳网络。 |
| id = id name = system_name Specifies the identification number（either 0 or 1）for the cluster system and the name that is returned by the hostname command（for example，storage0）. powerSerialPort = serial_port | 为电子开关所连接的串行口（如果有的话）指定设备专用文件（如/dev/ttyS0）。 |
| powerSwitchType = power_switch | 指定电子开关的类型，如 RPS10、APC 或 None。 |

在 config 文件中还有一个[sg]部分：

[sg] device0 = sg_device_name 用于为共享磁盘指定一个 sg 设备名称。

当您添加集群服务时，您所指定的服务信息将被加入一些字段，这些字段位于数据库中的[service]（服务）部分。下面是对集群服务字段的描述：

**表 2.**

| | |
|---|---|
| start service0 name = service_name disabled = yes_or_no userScript = path_name | 指定服务的名称（无论该服务在创建后是否被禁用）和用于启动和停止服务的任何脚本的完整路径名称。 |
| preferredNode = member_name relocateOnPreferredNodeBoot = yes_or_no | 指定您要在其上运行服务的系统的名称，并确定当该系统重启并加入集群之后是否仍重新定位到该系统。 |
| start servicecheck0 checkScript = path_name checkInterval = time checkTimeout = time maxErrorCount = number end servicecheck0 | 指定服务检查特性中需要用到的服务检查脚本（如果有的话）、检查时间间隔、检查超时和最多错误次数。 |
| start network0 ipAddress = aaa.bbb.ccc.ddd netmask = aaa.bbb.ccc.ddd broadcast = aaa.bbb.ccc.ddd end network0 | 指定服务使用的 IP 地址（如果有的话）、相应的子网掩码和广播地址。注意，您可以为一个服务指定多个 IP 地址。 |
| start device0 name = device_file | 如果有的话，指定服务中使用的专用设备文件（如：/dev/sda1）。注意，您可以为一个服务指定多个设备文件。 |
| start mount name = mount_point fstype = file_system_type options = mount_options forceUnmount = yes_or_no | 指定设备的加载点（如果有的话）、文件系统类型、加载选项，并指定加载点是否允许强行卸载。 |
| owner = user_name group = group_name mode = access_mode end device0 end service0 | 指定设备所有者、设备所属的组和设备接入模式。 |

## B.5 磁盘镜像配置

GreatTurbo Cluster server 10 使用 drbd 作为磁盘镜像工具。请访问 www.drbd.org，以得到到更多参考信息。GreatTurbo Cluster Server 10 支持 Linux kernel2.4 和 2.6 版本的 drbd。

### B.5.1 介绍

- 什么是 DRBD？

  DRBD 是由内核模块和相关脚本而构成，用以构建高可用性的集群。其实现方式是通过网络来镜像整个设备。您可以把它看作是一种网络 RAID。

- drbd 的应用范围是什么？除此之外，创建高可用性集群还需要什么？

  Drbd 负责接收数据，把数据写到本地磁盘，然后发送给另一个主机。另一个主机再将数据存到自己的磁盘中。

  其他所需的组件有集群成员服务，如 GreatTurbo Cluster Server 10 或心跳连接，以及一些能在块设备上运行的应用程序。

  例如：

  - 裸 I/O
  - 文件系统及 fsck
  - 具有恢复能力的数据库。

- 它是如何工作的？

  每个设备（drbd 提供了不止一个设备）都有一个状态，可能是‘主’状态或‘辅助’状态。在带有主要设备的节点上，应用程序应能运行和访问设备（/dev/nbX 或/dev/drbdX）。每次写入都会发往本地低层块设备和带有‘辅助’状态设备的节点中。次要设备只能简单地把数据写入它的低层块设备上。读取数据通常在本地进行。

  如果主要节点发生故障，心跳将会把辅助设备转换到主状态，并启动其上的应用程序。（如果您将它和无日志 FS 一起使用，则需要运行 fsck）。

如果发生故障的节点恢复工作，它就会成为新的辅助节点，而且必须使自己的内容与主节点的内容保持同步。当然，这些操作不会干扰到后台的服务。

- drbd 同现在的 HA 集群有什么关系？

  大部分现行高可用性集群（如：惠普、康柏等等）使用的是共享存储设备，因此存储设备连接多个节点（用共享的 SCSI 总线或光纤通道就可以做到）。

  Drbd 也可以作为一个共享的设备，但是它并不需要任何不常见的硬件。它在 IP 网络中运行，而且在价格上 IP 网络要比专用的存储网络经济的多。

  目前，drbd 每次只允许对一个节点进行读写访问，这对于通常的故障切换高可用性集群来讲已经足够用了。以后的版本将支持两个节点进行读写存取。这很有用，比如对 GFS 来讲就是如此。

### B.5.2 兼容性

Drbd 可以在 ide、SCSI 分区和整个驱动器之上运行，但不能在回路模块设备上运行。（如果您硬要这样做，它就会发生死锁）。

Drbd 也不能在回送网络设备中运行。（因为它同样会发生死锁：所有请求都会被发送设备占用，发送流程也会阻塞在 sock_sendmsg（）中。有时，接收线程正从网络中提取数据块，并试图把它放在高速缓存器中；但系统却要把一些数据块从高速缓存器中取到磁盘中。这种情况往往会在接收器的环境下发生，因为所有的请求都已经被接收器块占用了。）

### B.5.3 Linux kernel2.4 版本 drbd 的配置文件

在 Linux kernel2.4 环境下，GreatTurbo Cluster Server 10 所用的 drbd 的配置文件是在配置 GreatTurbo HA 时自动生成的，所以用户不用手动编辑 drbd 的配置文件。

Drbd 的配置文件的存放路径为：/etc/drbd.conf

下面为 drbd 的配置文件的一个例子：

```
resource drbd0 {
protocol=C
fsck-cmd=fsck.ext2 -p -y

on thost1 {
device=/dev/nb0
disk=/dev/hda7
address=10.1.1.31
port=7789
   }

on thost2 {
device=/dev/nb0
disk=/dev/hda7
address=10.1.1.32
port=7789
   }
}
```

下面对例子中的配置文件的内容进行说明：

| | |
|---|---|
| Resource drbd0 | 指定 drbd 资源名，每一个 resource 对应一个 drbd 设备。 |
| protocol=C | 指定 drbd 的协议。<br>Drbd 有 3 种协议：<br>A：数据一旦写入磁盘并发送到网络中就认为完成了写入操作。<br>B：收到对方的接收确认就认为完成了写 |

| | |
|---|---|
| | 入操作。<br>C：收到对方的写入确认就认为完成了写入操作。<br>在 GreatTurbo HA 中，默认采用最可靠的 C 协议。 |
| fsck-cmd=fsck.ext2 -p -y | 指定 fsck 的命令。 |
| On thost1 | 指定节点的机器名。 |
| device=/dev/nb0 | 指定 drbd 设备名，两个节点的 drbd 设备名要对应一致。 |
| disk=/dev/hda7 | 指定 drbd 设备所对应的物理磁盘分区。 |
| address=10.1.1.31 | 指定 drbd 设备所使用的网络 IP 地址。 |
| port=7789 | 指定 drbd 设备设备通讯所使用的端口，同一个 drbd 设备在两个节点上的端口要相同。 |

以上的例子为 GreatTurbo HA 自动生成的 drbd 配置文件，要了解 drbd 配置文件的更多信息，请参照 man drbd.conf。

## B.5.4 Linux kernel2.6 版本 drbd 的配置文件

在 Linux kernel2.6 环境下，用户需要手动编辑 GreatTurbo Cluster Server 10 所用的 drbd 的配置文件。

GreatTurbo Cluster Server 10 安装后，会生成一个默认的 Drbd 的配置文件：/etc/drbd.conf，用户只要编辑这个文件即可，然后将编辑好的配置文件拷贝到对方节点。

下面为 drbd 的配置文件的一个例子：

```
resource drbd0 {
  protocol C;
```

```
  startup {
     wfc-timeout    30;
     degr-wfc-timeout 60;
  }

  syncer {
     rate 600M;
     group 0;
  }

  on hostname1 {
     device         /dev/drbd0;
     disk           /dev/hda6;
     address        192.168.0.1:7788;
     meta-disk    internal;
  }

  on hostname2 {
     device       /dev/drbd0;
     disk         /dev/hda6;
     address      192.168.0.2:7788;
     meta-disk internal;
  }
}
```

下面对例子中的配置文件的内容进行说明：

| | |
|---|---|
| resource drbd0 | 指定 drbd 资源名，每一个 resource 对应一个 drbd 设备。 |

| | |
|---|---|
| Protocol C; | 指定 drbd 的协议。<br>Drbd 有 3 种协议：<br>A：数据一旦写入磁盘并发送到网络中就认为完成了写入操作。<br>B：收到对方的接收确认就认为完成了写入操作。<br>C：收到对方的写入确认就认为完成了写入操作。<br>在 GreatTurbo HA 中，默认采用最可靠的 C 协议。 |
| Startup | 指定 drbd 启动时的对数。 |
| wfc-timeout 30; | 启动后等待连接的 timeout。 |
| degr-wfc-timeout 60; | 单节点情况下重启后等待连接的 timeout。 |
| Syncer | 指定 drbd 同步时的参数。 |
| rate 600M; | 指定 drbd 同步时网络带宽的上限。 |
| group 0; | 指定 drbd 同步时的组，同一个组的 drbd 将并行进行同步。如果 drbd 设备位于不同的物理磁盘，请指定不同的组。 |
| on hostname1 | 指定节点的机器名。 |
| Device /dev/drbd0; | 指定 drbd 设备名，两个节点的 drbd 设备名要对应一致。 |
| disk /dev/hda6; | 指定 drbd 设备所对应的物理磁盘分区。 |
| address 192.168.0.1:7788; | 指定 drbd 设备所使用的网络 IP 地址和端口号，同一个 drbd 设备在两个节点上的 |

| | |
|---|---|
| | 端口要相同。 |
| `meta-disk  internal;` | 指定 drbd 的 meta-data 的存储位置。<br>Internal: 表示物理磁盘的最后 128MB 空间用来存储 drbd 设备的 meta-data。<br>还可以单独指定 meta-disk，例如指定 /dev/hde6 为 meta-disk ： meta-disk /dev/hde6[0];表示/dev/hde6 的前 128MB 空间用来存储 drbd 设备的 meta-data。 |

以上的例子为 GreatTurbo HA 所使用的默认 drbd 配置文件，要了解 drbd 配置文件的更多信息，请参照 man drbd.conf。

## B.5.5 drbd 的启动和停止

Drbd 的配置文件/etc/drbd.conf 准备好后，可以利用/etc/init.d/drbd 脚本启动和停止 drbd。

启动 drbd: /etc/init.d/drbd start

停止 drbd: /etc/init.d/drbd stop

另外，drbd 还提供了命令 drbdsetup 来配置 drbd，详细可以参照 man drbdsetup。

## B.5.6 Linux kernel2.4 版本/proc/drbd

Drbd 的状态可以通过/proc/drbd 文件来查看，GreatTurbo Cluster Server 10 所用的 Linux kernel2.4 版本的 drbd /proc 文件为：

```
[root@test1 root]# cat /proc/drbd

version: 0.6.8 (api:63/proto:62)
```

```
0: cs:Connected st:Primary/Secondary ns:104397 nr:0 dw:9 dr:104521 pe:0 ua:0
```

下面对/proc/drbd 文件的内容进行说明：

| Field | 说明 | 值 |
|---|---|---|
| Cs | 连接状态 | o Connected：一切正常<br>o Unconfigured：过渡状态，设备在等待配置(insmod 后的状态)<br>o StandAlone：过渡状态，配置完 disk 没配 net。<br>o Unconnected：连接模块时的过渡状态，或连接失败时的状态<br>o Timeout：发送数据包 timeout<br>o BrokenPipe：发送数据包错误<br>o WFConnection：连接之前的状态<br>o WFReportParams：发送参数给对方<br>o SyncPaused：等待高优先级的 group 先 sync 完<br>o SyncingAll：正将主节点的所有模块复制到次级节点上<br>o SyncingQuick：通过复制已被更新的模块（在次级节点短暂离开集群的情况下）来更新次级节点 |
| St | 状态 | 可能的值为：<br>o Primary/Secondary<br>o Primary/Unknown<br>o Secondary/Secondary<br>o Secondary/Unknown |
| Ns | 发送的数据 | KB |
| Nr | 接收的数据 | KB |
| Dw | 磁盘写入的数据 | KB |
| Dr | 磁盘读取的数据 | KB |
| Of | 运行中(过时)的请求 | 请求数 |
| Pe | 待解决的请求 | 请求数 |
| Ua | 未答复的请求 | 请求数 |

## B.5.7 Linux kernel2.6 版本/proc/drbd

Drbd 的状态可以通过/proc/drbd 文件来查看，GreatTurbo Cluster Server 10 所用的 Linux kernel2.6 版本的 drbd /proc 文件为：

```
[root@test1 root]#    cat /proc/drbd
```

```
version: 0.7.11 (api:77/proto:74)
SVN Revision: 1912M build by root@qa3-127, 2005-08-25 10:20:28
 0: cs:Connected st:Primary/Secondary ld:Consistent
    ns:20387 nr:366908 dw:387295 dr:14614 al:47 bm:139 lo:0 pe:0 ua:0 ap:0
```

下面对/proc/drbd 文件的内容进行说明：

| Field | 说明 | 值 |
|---|---|---|
| Cs | 连接状态 | o Connected：一切正常<br>o Unconfigured：过渡状态，设备在等待配置(insmod 后的状态)<br>o StandAlone：过渡状态，配置完 disk 没配 net。<br>o Unconnected：连接模块时的过渡状态，或连接失败时的状态<br>o Timeout：发送数据包 timeout<br>o BrokenPipe：发送数据包错误<br>o NetworkFailure：网络错误<br>o WFConnection：连接之前的状态<br>o WFReportParams：发送参数给对方<br>o SkippedSyncS：应当为同步的源端，但用户选择不同步<br>o SkippedSyncT：应当为同步的目的端，但用户选择不同步<br>o WFBitMapS：过渡状态，同步源端同步前等待 bitmap<br>o WFBitMapT：过渡状态，同步目的端同步前等待 bitmap<br>o SyncSource：同步的源端<br>o SyncTarget：同步的目的端<br>o PausedSyncS：同步源端暂停<br>o PausedSyncT：同步目的端暂停 |
| St | 状态 | 可能的值为下列值的组合：<br>o Primary<br>o Secondary<br>o Unknown |
| ld | 物理磁盘的状态 | o Consistent：数据一致<br>o Inconsistent：数据不一致 |
| Ns | 发送的数据 | KB |
| Nr | 接收的数据 | KB |
| Dw | 磁盘写入的数据 | KB |

| | | |
|---|---|---|
| Dr | 磁盘读取的数据 | KB |
| Al | 访问 log 的数目 | 个数 |
| Bm | 写 bitmap 的数目 | 个数 |
| Lo | 等待本地磁盘来标记操作完成的请求 | 请求数 |
| pe | Pending(等待应答的请求) | 请求数 |
| ua | 没有应答的请求 | 请求数 |
| ap | 等待结束的请求 | 请求数 |

### B.5.8 有关 drbd 的 Q&A

- 问：我能加次级设备吗（至少只读）？

  答：DRBD 不允许加载次级设备。

- 问： DRBD 能使用两个容量大小不同的设备吗？

  答： 一般情况下可以，但有些问题需要注意：

  本地 DRBD 使用的是配置的磁盘容量，与物理容量相等。如果没有给出，则将被设置为物理容量。连接时，设备容量将设置为两个节点中最小容量。

  如果缺少常识的话，您可能会碰到一些问题：如果您先是在一个节点上使用 drbd ，而且没有配置好磁盘容量，之后又连接了一个容量较小的设备。这时，drbd 设备容量在运行时就会变小。在系统记录里，您会发现一条信息提示 “ your size hint is bogus,please change to some value ”（您的容量信息不真实，请更改）。这样一来就会让设备顶层的文件系统造成混淆。

  因此，如果您的设备容量不同，请明确地为 DRBD 设置所使用的容量。

- 问： XFS 能和 DRBD 一起使用吗？

  答：Linux kernel2.6 版本的 DRBD 支持 XFS。

- 问：当我试着加载 drbd 模块时，遇到了下面的问题：

  compiled for kernel version 2.4.18-4GB while this kernel is version 2.4.18-64GB-SMP.

- 答：您的实际内核与要在其上构建 drbd 的内核的.config 不一致。请更换成与 drbd 相匹配的内核，或与拓林思公司联系获取与您的内核相匹配的 drbd。

## B.6 调整 Oracle 服务

故障切换后，Oracle 数据库的恢复时间与未完成的交易数量和数据库的大小成正比。以下参数控制数据库的恢复时间：

- LOG_CHECKPOINT_TIMEOUT
- LOG_CHECKPOINT_INTERVAL
- FAST_START_IO_TARGET
- REDO_LOG_FILE_SIZES

如果想把恢复时间降至最低，则将上述参数设置为相对较低的值。注意，值过低将对性能产生不利影响。 您可能需要重新设置，以找到一个最理想的值。

Oracle 还提供了其它控制数据库交易重试和重试延迟时间的调整参数。这些值必须足够大，以包含您的环境中进行故障切换的时间。这样可确保故障切换对数据库客户机应用程序是透明的，并且无需程序重新连接。